

Meanings of Diversity in the *Chronicle of Higher Education*, 2019–25: A Hermeneutic Text Mining Approach

Kyle Conway, Sarah Donnelly, Madison Lagasse Brockbank

2026-01-12

Abstract

In the past decade, few policies have been politicized in U.S. higher education as much as those meant to encourage diversity among students and professors. DEI (diversity, equity, and inclusion) policies have become a favourite target among politicians appealing to voters who see them as a symptom of a zero-sum game where people not perceived as “diverse” are valued less than others. This article examines 118 articles about diversity published in the *Chronicle of Higher Education* between 2019 and 2025. It uses text mining tools (specifically, sentiment, correspondence, and correlation analysis) to map changes in the ways the words related to diversity, in particular *DEI* and *woke*, were used. It then employs hermeneutic techniques, drawing on the work of Paul Ricoeur, to interpret the patterns identified through text mining. It shows points of continuity, resulting from the ways speakers respond to past discourse, and change, resulting from the ways they anticipate others’ responses and change their behaviour accordingly. Ultimately, this article documents an evolution in meanings of diversity characterized by growing pessimism in response to national and international politics and conflict.

Keywords

DEI; Charles Peirce; hermeneutics; higher education; text mining; woke

Introduction: The changing meanings of diversity

In the past decade, few policies have been politicized in U.S. higher education as much as those meant to encourage diversity among students and professors. DEI (diversity, equity, and inclusion) policies have become a favourite target among politicians appealing to voters who see them as a symptom of a zero-sum game where people not perceived as “diverse” are valued less than others, despite the stated intentions of the policies’ architects.

One striking feature of the controversies surrounding these policies is the way that usage of words referring to diversity has evolved. Such words include *diversity* itself, as well as *woke* and *DEI*. At the most basic level, their use has increased. For example, use of *woke*, referring broadly to people’s awareness of the stakes of diversity, began to increase as early as 2000, while use of *inclusion* and *diversity* picked up starting in 2015 (figure 1). The frequency of *equity*, too, has increased, although it peaked in 2008 at the height of that decade’s recession, referring presumably to the equity people lost when real estate prices dropped.

In this article, we combine text mining and hermeneutics to describe how the use of such words changed between 2019 and 2025, focusing on articles published in the *Chronicle of Higher Education*, which describes

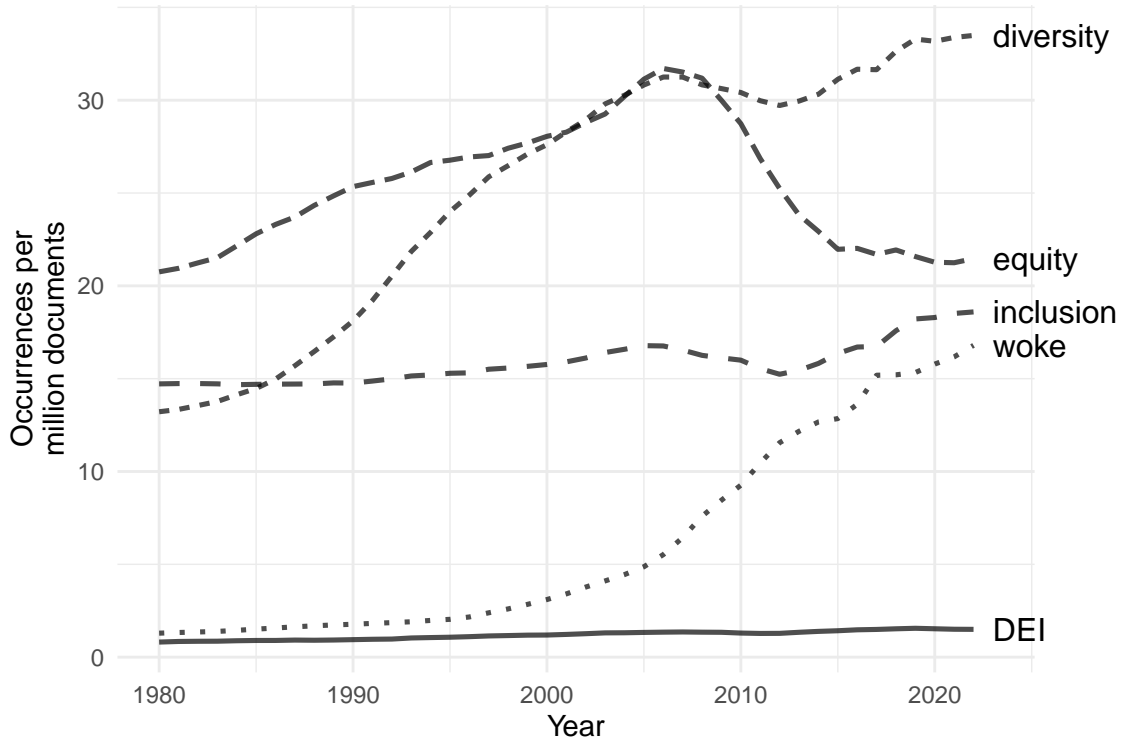


Figure 1: N-grams showing the frequency of usage of the words *diversity*, *equity*, *inclusion*, *woke*, and *DEI*, 1980–2022. Data source: Google N-grams.

itself as “academe’s most trusted resource for independent journalism, career development, and forward-looking intelligence” (*Chronicle of Higher Education* 2025a). We chose the *Chronicle of Higher Education* for its reach and prominence within U.S. higher education: as of September 2025, it had 27,000 individual subscribers, 1,800 institutional subscribers, 2.1 million monthly site visits, and 1 million newsletter subscribers (*Chronicle of Higher Education* 2025b).

Broadly speaking, we investigate the evolution of words’ semantic networks, treating meaning as a function of word use in identifiable contexts. We use three tools to cluster words, namely sentiment, correspondence, and correlation analysis, which allow us to identify changing patterns of words used together. We proceed, first, by describing the rhetorical dimensions of words such as *DEI* and especially *woke*, whose capaciousness allows speakers to mobilize a range of contradictory meanings. We then turn to the knotty problem of the relationship between text mining, which effects an epistemological break by treating texts as unordered collections of words, and meaning, from which the patterns identified by text mining appear divorced. We suggest that these patterns can function as hermeneutic “guesses,” following Paul Ricoeur (1981), for whom the interpretation of a text consists in a cycle of guessing, testing, and refinement. From there, we turn to the tools themselves, describing how they function and what they reveal. Patterns emerge from this analysis, first, in words’ emotional valence observed over time, and second, in thematic clusters associated with specific periods and events. Using these patterns as a hermeneutic starting point, we read them back into our corpus, examining specific articles where they appear. We observe points of both continuity and change: over the course of the six years we examine, discussions of diversity grew more pessimistic, covering a wide range of topics, from admissions, hiring, and promotion in 2019–23, to the conflict between Hamas and Israel in 2024, to changes in federal antidiscrimination regulations in 2025. We conclude by producing a map that represents these evolving sets of associations.

Semantic and rhetorical dimensions

As we write below, our focus is on two specific words, *DEI* and *woke*, as they relate to ideas of diversity. Although DEI programs have been widely discussed in fields such as education and social work, only rarely have scholars examined the rhetorical dimensions of the word *DEI* itself. Some have defined the word by identifying the policies associated with it (Iyer 2022; Mogilski et al. 2025), or by speaking with practitioners to understand how they interpret its component parts, *diversity*, *equity*, and *inclusion* (Reisch and Jani 2025). How speakers mobilize the term rhetorically, however, remains under-explored.

In contrast, the rhetorical dimensions of *woke*, which has proven more polarizing, have been widely discussed. The term is what W.B. Gallie calls an “essentially contested concept,” “the proper use of which inevitably involves endless disputes about [its] proper uses on the part of [its] users” (Gallie 1956, 169). Such concepts contain multiple, complex parts that seem to orbit around a fixed core, but beyond the idea that such a core exists, or that a concept is internally complex, there is little agreement. According to the idea’s supporters, to be woke is “to wake up from a false sense of security that [people] are safe or free in the bodies they occupy,” an act that encourages “self-determination, citizenship formation, and engaged activism” (Cherry-McDaniel 2017, 42). According to its critics, the term is associated with cancel culture, or the deplatforming of speakers who contradict a perceived liberal dogma and impute guilt to members of racial majorities, especially white people, based on their race (Fon 2024). Or, going still further, it is a new religion, replete with its own clergy (authors of popular books about racial justice) promoting ideas of sin and redemption (white privilege and its renunciation) in ways that parallel Christian evangelism (McWhorter 2021).

Indeed, the evolution of uses of *woke* is a valuable index for evolving attitudes toward – and political uses of – diversity more broadly. Use of the term *woke* originated in Black communities in the United States (Cherry-McDaniel 2017). Bart Cammaerts traces it back to Lead Belly’s 1938 protest song “Scottsboro Boys,” where “staying woke explicitly referred to the need for Afro-Americans to be acutely aware and conscious of the dangers and threats that were inherent to a white-dominated racist America” (Cammaerts 2022, 734). The contested nature of *woke* is observable in the changes the term has undergone in the last decade. It gained prominence during the protests following the death of Michael Brown at the hands of a police officer in Ferguson, Missouri, in 2014. From then to 2019, it operated largely within a semantic field of ideas related to citizenship and empowerment, especially within Black communities (Cherry-McDaniel 2017). This was true outside of the United States, too: the first time the word *woke* appeared in the *Hansard* of the Canadian House of Commons was 2019, when a Member of Parliament (MP) used it to draw attention to Black History Month (McCurdy, Clark, and Cammaerts 2025, 920). During this time, public opinion generally supported social movements falling under the *woke* label, such as Black Live Matters, itself growing out of the Ferguson protests. Protests following the 2020 killing of George Floyd attracted an increasingly diverse set of participants (Roberts 2021). White progressives in particular adopted the term *woke* to “encompass broader societal injustices related to feminist, LGBTQ, and intersectional concerns” (McCurdy, Clark, and Cammaerts 2025, 914).

Patrick McCurdy, Kaitlin Clark, and Bart Cammaerts (2025) identify this expansion as the first in a series of semantic shifts. As the term and the ideas it represented became more popular, they also provoked a backlash, marking a second shift. For instance, John McWhorter wrote in his book *Woke Racism* that the growing dogmatism around wokeness, especially its adherents’ insistence (as he saw it) that anything short of absolute allegiance to antiracism was itself racist, “forces us to render a great deal of our public discussion of urgent issues in double-talk any ten-year-old can see through” (McWhorter 2021, chap. 1). In the early 2020s, politicians used wokism as a foil, the basis of an accusation against their opponents. Their strategy consisted in reversing the accusation of discrimination made by defenders of social justice, arguing that their accusers were the ones who were “irrational, exaggerated, moralizing, intolerant, polarizing, radical, and totalitarian,” while they themselves were the ones who upheld “common sense, rationality, respect for nature and tradition, enlightenment and freedom of speech” (Dhoest and Paulussen 2025, 1930; see also Asen 2024). Their most common tactic consisted in manufacturing a crisis or moral panic: in contrast to their “woke” accusers, they said, they were interested in “rational debate and reasonable disagreement” (Isaac Kamola, in Fekete 2022, 51). Such accusations reached a “crescendo” in about 2023, marking a third shift, where politicians came to associate wokeness not just with irrationality or the suppression of free speech,

but with an authoritarianism outside the bounds of – or even opposed to – the nation whose traditions they claimed to uphold. In the Canadian context, for instance, “As the use of woke as a pejorative term became more entrenched, Conservative MPs began constructing a more abstract and menacing image of wokeness as an authoritarian Other, inverting the roles of oppressed and oppressor” (McCurdy, Clark, and Cammaerts 2025, 925).

Thus, over the course of a decade, the semantic field in which the term *woke* circulated changed. Originally associated with ideas of citizenship and justice within Black communities, it expanded to include gender and sexual orientation. Politicians came to use it to accuse their opponents of advancing values diametrically opposed to the fairness and justice they claimed to support.

This increasingly adversarial attitude, we will see, was observable in the articles about diversity published in the *Chronicle of Higher Education*, too.

Text mining and hermeneutics

Our method uses algorithms written in the language R (although the underlying logic could be expressed in other languages) to group words into clusters that suggest relationships between ideas (R Core Team 2025). As we write in the introduction, it employs three techniques. The first is sentiment analysis, which measures the ratio of words with positive connotations to words with negative connotations. The second is correspondence analysis, which reveals relationships between articles (grouped by year) and the words they contain. The third is correlation analysis, which calculates the likelihood that word *X* will appear in proximity to word *Y*. These techniques allow us to establish patterns broadly across our corpus, a form of distant reading (Moretti 2000). These patterns then guide us in a more conventional mode of analysis where we examine individual texts within our corpus, reading clusters back into articles themselves to determine which dimensions of meaning they reflect (and which they do not) (Aledavood 2024).

An epistemological break

Our approach runs into an immediate problem: treating a corpus of texts – or, more accurately, the patterns of words within texts that make up a corpus – as data presupposes data as “*a priori* and collectible” things “that one gathers” (Markham 2013, sec. 1). The “bag of words” approach, which we adopt, further simplifies the complexity of language by quantifying occurrences of words. It transforms texts into sets of words whose contents can be compared to other sets, themselves transformed from other texts, accounting for word frequency but not word order (Leetaru 2012, 69). The idea that the connections traced among words reflect the structure of associations they evoke – their meaning, as we argue below – must be carefully justified, given the obvious objections: “language,” Zellig Harris wrote seven decades ago, “is not merely a bag of words but a tool with particular properties which have been fashioned in the course of its use” (1954, 156). Imputing sentiment or calculating degrees of word correlation induces an epistemological break, bracketing off meaning and leading to a “project of pure description of discursive events” (Scholz 2019, 127–8).

In effect, our approach demonstrates that “all quantitative models of language are wrong,” as Justin Grimmer and Brandon M. Stewart (2013, 271) observe. But does it also show how “some are useful” (Grimmer and Stewart 2013, 270)?

Questions of semantics

Our concern is with words’ semantic dimensions. Our approach derives from the work of Charles Peirce, who describes signs as consisting of three components. First is the *representamen*, or “something which stands to somebody for something in some respect or capacity” (Peirce 1940, 99). Representamens can take many forms, such as an image, a sound, or a word. Second is the *object*, or the thing in the world to which the representamen refers, and third is the *interpretant*, or “an equivalent sign, or perhaps a more developed sign”

evoked by the representamen (Peirce 1940, 99). Interpretants, too, can take many forms, such as images, sounds, words, and even emotions. Because interpretants are signs, they can also act as representamens, which is to say, they can evoke further associations – further interpretants – which can evoke still others. In other words, signs function by linking representamens, objects, and interpretants in a network of associations, which themselves are embedded in even larger networks, in ways suggested (in simplified form) by figure 2.

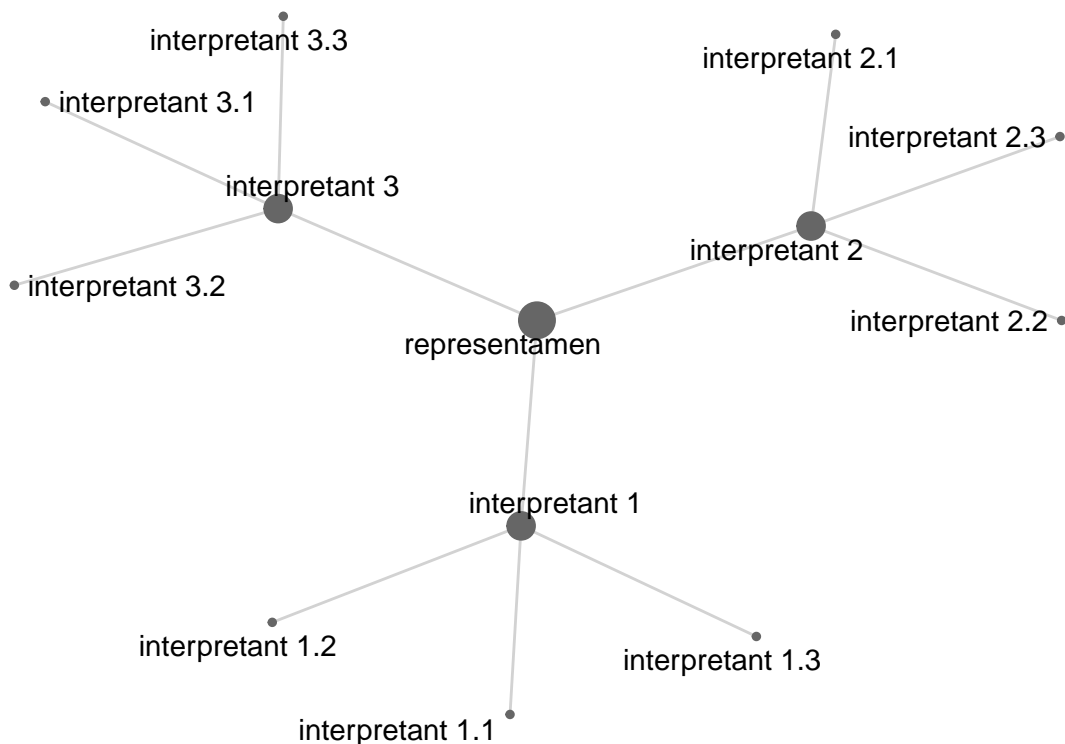


Figure 2: A sign as theorized by Peirce, including a representamen and its first- and second-order interpretants.

To overcome this epistemological break, we return to research conducted in the 1950s. Our claim that word clusters provide clues about texts’ meaning derives in part from the relation between predictability and word function. Early research in the cybernetic tradition, for instance, demonstrated that words appear in a statistically predictable order, where “past events influence present probabilities” (Weaver 1949, 15). For instance, the article *the* is likely to be followed by a noun (or, in some cases, an adjective modifying a noun, or in rarer cases, an adverb modifying an adjective modifying a noun). It is unlikely to be followed by another article or a verb.

Similar patterns can be observed with respect to word choice. For instance, Harris observes that “distributional relations [...] tell us something about the occurrence of elements and [...] correlate with some aspect of meaning” (1954, 156). The words *oculist* and *eye doctor*, to borrow his example, refer to the same occupation, but they are used in different contexts, reflecting meaning as a relationship defined by context observable in the words’ professional or familiar registers. J.R. Firth uses such observations to make a larger point about collocations, or pairs of words that appear in proximity to each other:

The habitual collocations in which words under study appear are quite simply the mere word accompaniment, the other word-material in which they are most commonly or most characteristically embedded. It can be safely stated that part of the “meaning” of *cows* can be indicated by such collocations as *They are milking the cows*, *Cows give milk*. The words *tigresses* or *lionesses* are not so collocated and are clearly separated in meaning at the *collocational level*. (Firth 1968, 180, original emphasis)

Or, as Firth says more pithily earlier in the same essay, “You shall know a word by the company it keeps” (1968, 179).

We take this observation further by observing that speakers orient themselves toward past and future speech acts. They orient themselves toward the past, first, by taking account of how the word they choose has been used before:

A word presents itself not as an item of vocabulary but as a word that has been used in a wide variety of utterances by co-speaker A, co-speaker B, co-speaker C and so on, and has been variously used in the speaker’s own utterances. (Vološinov 1986, 70)

Those past uses are interpretants of a speaker’s chosen word. Speakers know that they share knowledge of these past uses with their listeners, although they also know that their respective sets of interpretants are unlikely to overlap completely. These networks of associations become material they can harness to build an argument, their goal being to persuade their listeners to arrange their interpretants in a similar way. More simply, a speaker’s goal is to evoke for their listener a chain of associations similar to their own.

Thus, speakers also orient themselves toward the future, anticipating the responses they are likely to evoke from their listener and adjusting accordingly:

When constructing my utterance, I try actively to determine [the response of the person I am talking to]. Moreover, I try to act in accordance with the response I anticipate, so this anticipated response, in turn, exerts an active influence on my utterance (I parry objections that I foresee, I make all kinds of provisos, and so forth). (Bakhtin 1986, 95)

We contend that one factor contributing to the patterns of collocation (which we measure as word correlation) described by Firth and Harris is this reactive and anticipatory process. Words appear together with observable frequency because speakers are trying to influence the chains of associations they evoke, prompting them to repeat words in observable patterns.

Hermeneutic validation

However, the calculations we perform say nothing about meaning as such. Our contention that word clusters reflect meaning needs a different type of warrant. For that, we turn to the hermeneutic approach elaborated by Paul Ricoeur (1981). Analysis here consists in moving between a text, which has a fixed form and is for Ricoeur an objective fact, and interpretation, which derives from a reader’s capacity for making meaning and represents a text’s subjective dimension (Ricoeur 1981, chaps. 1 and 4; see also Conway 2023, chap. 2; Piper 2015). The relationship between a text’s objective and subjective (or interpretive) dimensions is mediated by a process of making and validating guesses (figure 3). Readers make guesses by observing “‘clues’ contained in the text itself” (Ricoeur 1981, 137), which allow them to formulate possible interpretations. They validate their guesses by assessing their degree of *congruence*: to choose between two interpretations of a text is to identify which is more probable, or which “takes account of the greatest number of facts furnished by the text [...] and] offers a qualitatively better convergence between the features which it takes into account” (Ricoeur 1981, 137). They also assess their guesses’ degree of *plenitude*, understood as the degree to which “all the connotations which are suitable [are] attributed” (Ricoeur 1981, 138, quoting Monroe Beardsley).

In our case, we are interpreting a corpus of articles published in the *Chronicle of Higher Education*, rather than an individual text, and our guesses derive from the clusters generated by our algorithms, as suggested in figure 3. The patterns we identify allow us to reconstruct writers’ arguments, but in ways that require verification. This is how we move past the algorithms’ epistemological limits. Sentiment analysis and semantic networks derived from correspondence and correlation analyses show broad patterns, such as changes in usage rates of words with positive or negative connotations, but they do not show what evidence writers use to make their claims, or how they justify their arguments. Nor do they show how individual writers situate themselves with respect to the conversation in which they are participating. By assessing the guesses we can derive from these broad patterns, moving back and forth between the text’s objective and interpretive dimensions, we can map out the complex networks of arguments and associations in which they circulate.

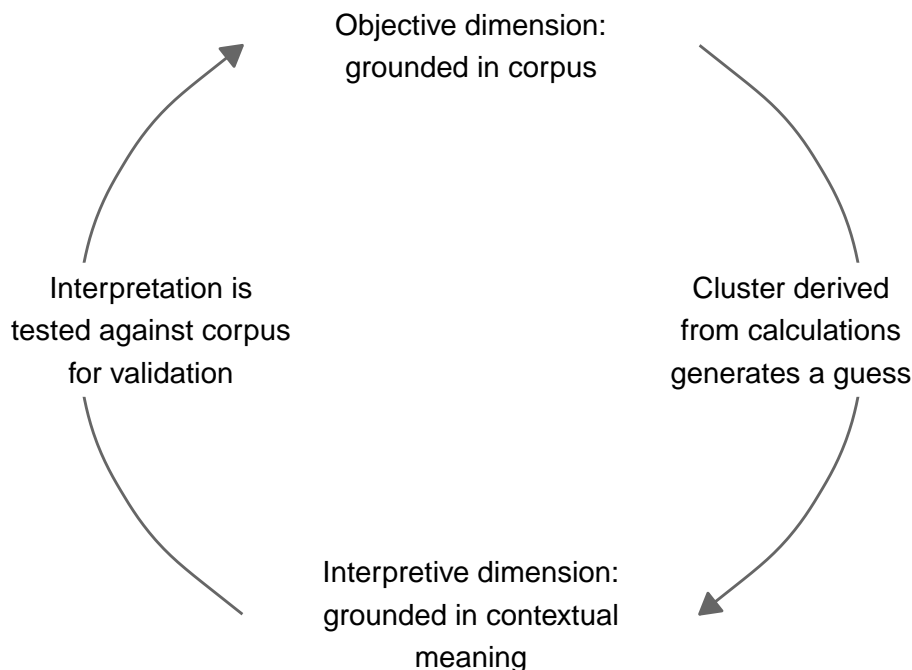


Figure 3: A hermeneutic circle where the objective and interpretive dimensions of a corpus are mediated through guesses and acts of validation.

Corpus construction

Our corpus consisted in articles published in the *Chronicle of Higher Education* between April 2019 and May 2025. We built it by performing two full-text searches on the Gale Academic OneFile database, first for articles containing both *DEI* and *diversity*, second for articles containing both *woke* and *diversity*. We performed these searches in February 2025 and again in August 2025. After removing false positives (including articles about Opus Dei and articles where *woke* was the past tense of *wake*), our corpus contained 118 articles. Using the `tidytext` and `textstem` packages in R (Silge and Robinson 2016; Rinker 2018), we tokenized and lemmatized the corpus; after removing stop words, using a list derived from the default provided by the `textstem` package, our corpus contained 103,248 lemmatized words.

The rate of articles published increased over the period we observed: 2 relevant articles were published in 2019, 2 in 2020, 7 in 2021, 7 in 2022, 46 in 2023, 33 in 2024, and 21 in 2025. Articles containing *DEI* outnumbered those containing *woke* (figure 4). (Note that for purposes of analysis, all letters are converted to lower case; hence, figures display *dei* instead of *DEI*. Proper nouns are also displayed without capital letters.)

The value of these articles as an object of study derives not from any putative representativeness – they cannot, in fact, stand in for a larger discourse – but instead from the investment that their authors and readers had in the outcome of the debates. They were reacting to and anticipating how other people used words related to diversity, and, given the broad circulation of the *Chronicle of Higher Education*, they were well placed to influence others.

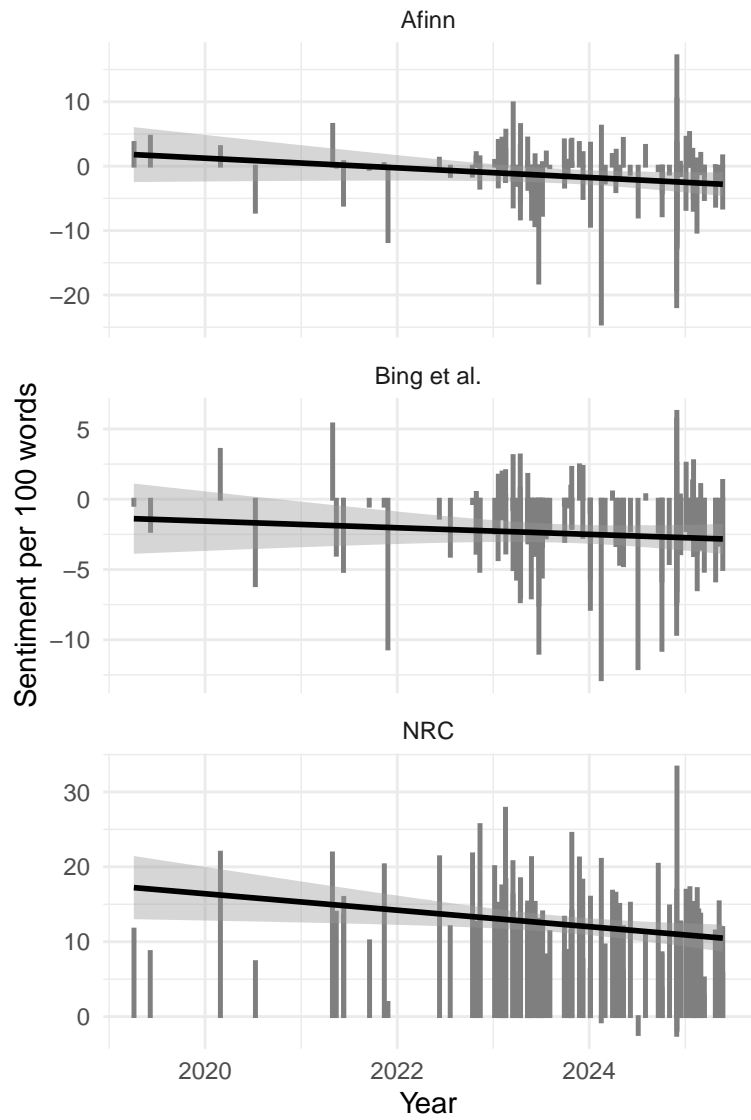


Figure 5: Sentiment analysis using the Afinn, Bing et al., and NRC lexicons, between April 2019 and May 2025.

idea of an interaction matrix from spatial data analysis, which uses such matrices to calculate the degree to which entities in space, represented as objects on a map, influence each other. Interaction matrices record which points are adjacent to each other and assign to those adjacencies a degree of attraction. The relative weight of each point is calculated by adjusting interaction measurements to make the total of each row equal to 1. The resulting sum of each column describes a point’s relative weight: the more points to which a point is close, and the more points to which those points are close (and so on down the line), the greater its relative weight (O’Sullivan and Unwin 2010, 47–9).

To determine the central words in our corpus, we treated words as adjacent if they appeared together in an article and defined the degree of attraction as the number of articles in which they appeared together. We calculated the degree of correspondence between the twenty-five most central words and each time period (figure 6).

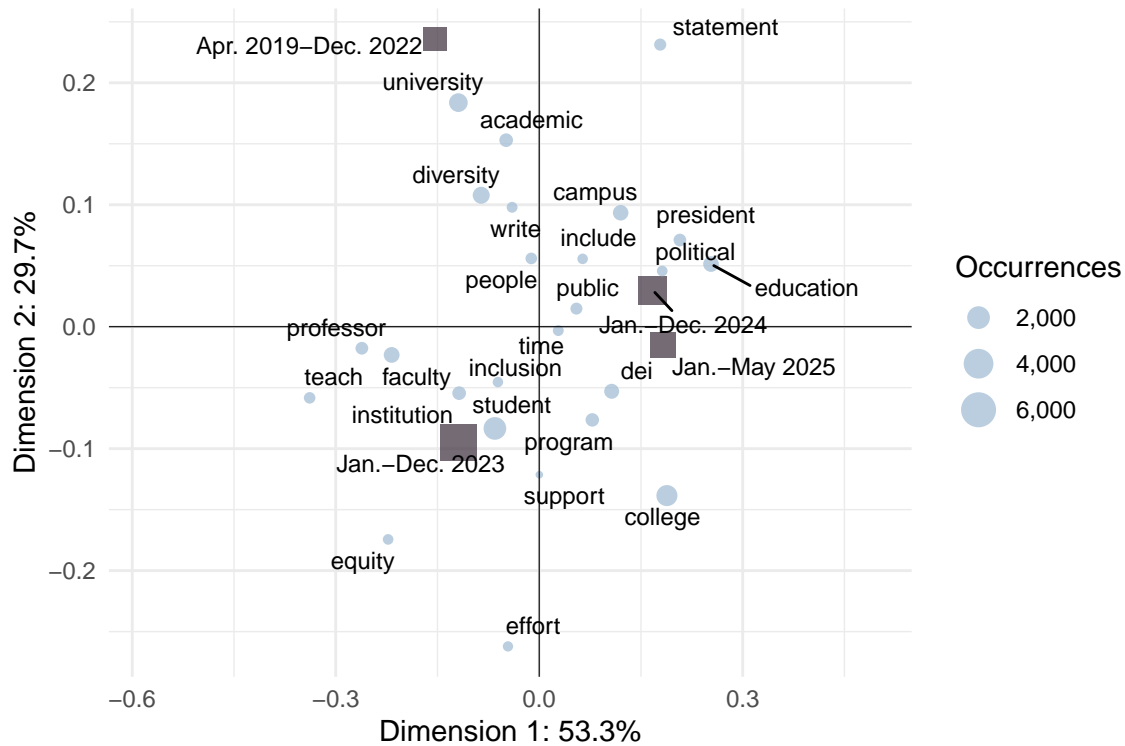


Figure 6: Correspondence between central words and *Chronicle of Higher Education* articles grouped by year.

Our second analysis revealed clearer clusters by identifying the words with the highest chi-squared residual values in relation to each period (figure 7). Correspondence analysis is in effect a re-expression of chi-squared values, and this approach identifies the pairs with the highest degrees of correspondence.

For our third analysis, we identified the words that were most distinct within each period by calculating their tf-idf (term frequency-inverse document frequency) values (Salton and Buckley 1988; Silge and Robinson 2017, chap. 3). These values are expressed as a ratio between the number of times a word appears in a given unit (in our case, a given period) and the number of units in which a word appears. Words that appear frequently in one unit but not in the others score higher than those that are evenly distributed across units. We calculated the degree of correspondence between the twenty-five words with the highest tf-idf values and each period (figure 8).

Correspondence analysis expresses relationships radially around the origin of the graph, where the x- and y-axes intersect. The more acute the angle between the vectors from the origin to a unit and to a word (a square and a circle in our maps), the greater the correspondence. In addition, points further from the origin

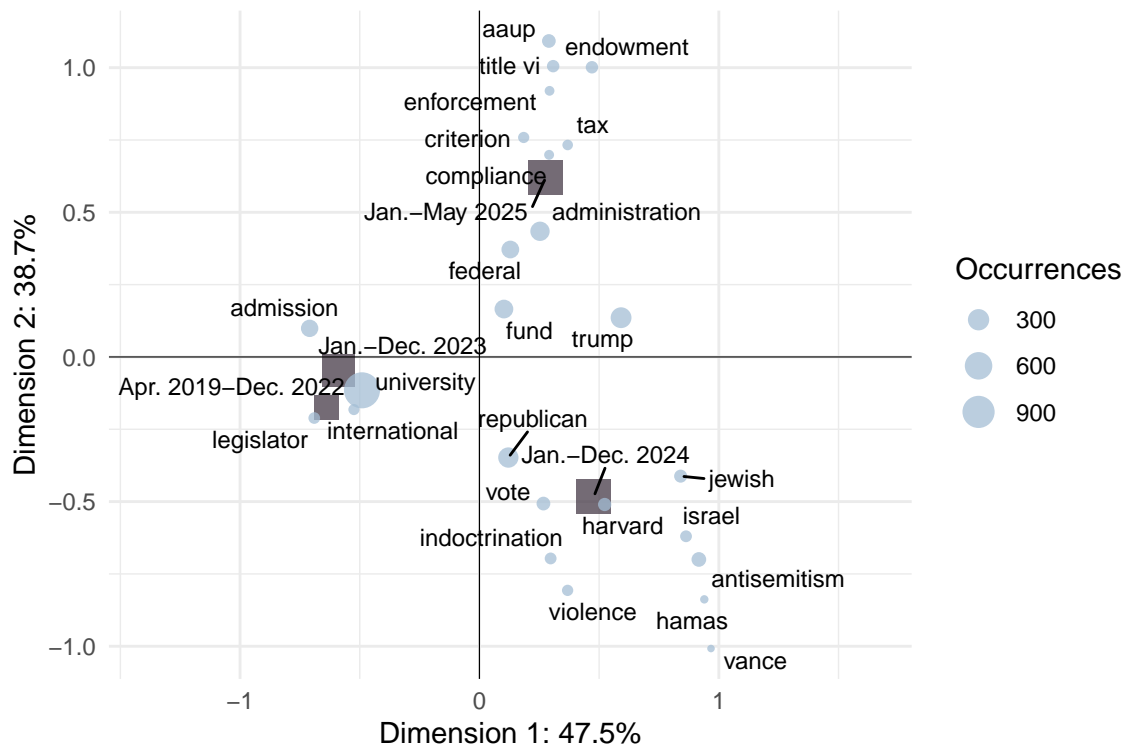


Figure 7: Correspondence between words with the greatest chi-squared values appearing at least 25 times and *Chronicle of Higher Education* articles grouped by year.

are more discriminating than closer points (Bock 2011). Two-dimensional graphs, such as figures 6 to 8, cannot capture the full variance of the relationships they describe; instead, they display the two dimensions accounting for the highest degrees of variance, indicated as percentages in the labels of the x- and y-axes.

Figures 6 to 8 reveal different dimensions of the debates about diversity in the *Chronicle of Higher Education* between 2019 and 2025. Figure 6 shows continuity between periods. Its points are clustered near the origin, with little distinction between periods, which is to be expected, given that it maps the most central words. Figure 7, on the other hand, shows three relatively distinct clusters: April 2019–December 2022 and January–December 2023; January–December 2024; and January–May 2025. Figure 8 reveals similar clusters, often including the same words as figure 7.

The difference between the number of occurrences of each term – words in figure 6 occur about 5 times as frequently as those in figure 7 and 10 times as those in figure 8 – also suggests that articles between periods were more similar than different. Those from which the words in figures 7 and 8 are drawn, while distinctive, constitute a proportionately small portion of the corpus.

Correlation analysis

So far, our analysis has related to congruence, which is concerned with content: what, collectively, were these articles about? What patterns emerged from the corpus as a whole? At a basic level, the clusters revealed by correspondence analysis were lists of words showing these patterns. If our search terms – *DEI* and *woke*, each paired with *diversity* – were representaments, correspondence analysis identified first-order interpretants, or words evoked by our search terms. It suggested three distinct periods, characterized both by a series of overarching themes and by concerns related to specific events. For the period from 2019 to 2023, marked by the Supreme Court’s decision banning the consideration of race in college admissions,

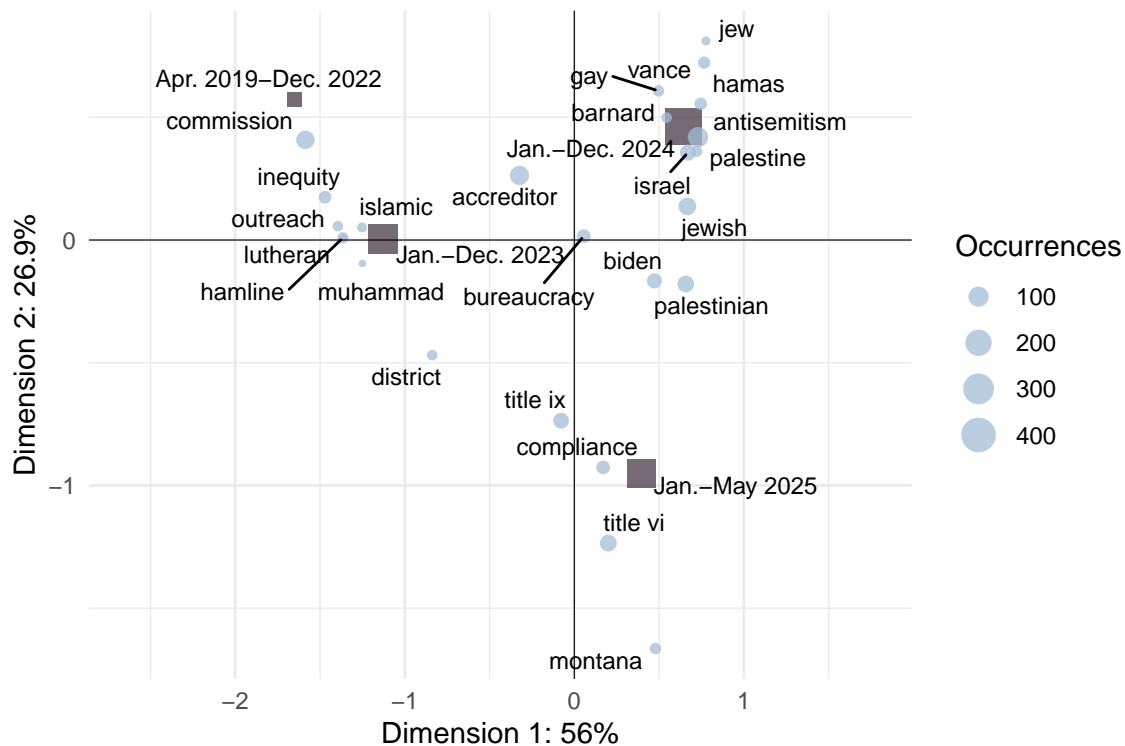


Figure 8: Correspondence between distinctive words, as determined by tf-idf values, and *Chronicle of Higher Education* articles grouped by year.

interpretants included words such as *university*, *admission*, *legislator*, and *outreach*. For 2024, they included words related to the conflict between Israel and Hamas, such as *attack*, *protest*, and *war*, and to Congress’s antisemitism hearings in late 2023. For 2025, they included words related to antidiscrimination legislation, in particular Title IX (banning sex-based discrimination) and Title VI (banning discrimination based on race, colour, or national origin). Correspondence analysis thus complemented our sentiment analysis, which suggested increasing levels of disagreement, marked by expressions of anxiety and even acrimony. To return to Ricoeur (1981), these observations constituted our initial guess, or our starting point for an account of the corpus’s different orders of meaning.

We turn now to plenitude, which is concerned with connections: what links did authors establish between ideas, both within individual articles and between articles and the larger social context? Correspondence analysis provided information to answer these questions, too. Specifically, it revealed connections between terms, in the chains of associations from the representamens to the first-order interpretants. To extend our analysis, we added a third measurement, that of correlation, to identify second-order interpretants, or ideas evoked by the words in the clusters around our initial search terms.

Correlation analysis measures the likelihood that two words will appear in proximity to each other. It derives from a contingency table counting how many times words X and Y appear together or not (table 1). The formula for calculating the correlation coefficient ϕ is

$$\phi = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

where A , B , C , and D are the values in table 1. Values for ϕ range between -1 and $+1$, negative values indicating that where one word appears, the other is *likely not to appear*, positive values indicating that where one word appears, the other is *likely to appear*. Absolute values of ϕ between $.10$ and $.29$ suggest a

small effect size, between .30 and .49 a medium effect size, and greater than .50 a large effect size (Wiedmaier 2018).

Table 1: Contingency table on which a correlation analysis is based.

	Contains word X	Does not contain word X
Contains word Y	A	B
Does not contain word Y	C	D

The articles in our corpus varied considerably in length, between 35 and 4,528 lemmatized words, after the stop words were removed. To account for this variation, and on the assumption that proximity indicates relatedness, we further divided each article into chunks of 200 consecutive lemmatized words.

To account for the plenitude of meaning, we returned to the articles themselves. We used correlation analysis to choose articles, creating maps by identifying pairs of first-order interpretants likely to appear together in the 200-word chunks into which we broke the corpus. We then read the articles containing each pair to see how authors used the words *DEI* and *woke* and their first- and second-order interpretants. Our goal was to describe the logics of connection: to whom were speakers responding? How did they position themselves with respect to others? What links did they draw between ideas?

2019–23: *Admission and university*

Congruence

For the first period, 2019–23, we mapped words related to *admission* and *university* (figure 9). Our map revealed two non-overlapping clusters, one around each word. The cluster around *university*, including words such as *political*, *faculty*, and *report*, suggests the ordinariness of debates about diversity measures, where the principal actors were professors and administrators working on a campus they shared. The cluster around *admission* suggests the nature and terms of the debates people were having. Many articles were about the Supreme Court’s decision (in June 2023) with respect to race-based admissions. Hence the words describing key points in the admissions process – *decision*, *applicant*, *enrollment* – as well as the categories universities used to measure diversity, including race (in *race*, *racial*, *minority*) and socio-economic status (in *low*, as in *low income*).

During this period, several authors offered definitions of key terms, such as *equity* (“an effort to ensure equal opportunity, not equal results”) and *inclusion* (“an effort to make everyone feel like they belong and are full and welcome members of the community”) (Ryan 2023, 42–3; see also Hogan and Sathy 2023). Overall, as the sentiment analysis above suggests, articles were positive, expressing support for the underlying assumption that diversity was an asset: “Many feel at a gut level what a majority of Supreme Court justices did not: Diversity matters in education; many young people have a compelling interest in it; and they can explain how it benefited them” (Hoover 2023, 23; see also Mangan 2023; for a contrary view, see Field 2023).

Plenitude

To get a sense of the relationship between interpretants characterizing this period, we extracted all articles that included both *university* and *admission* and explored the range of perspectives adopted by authors and the people they quoted, in particular faculty and students. The Supreme Court’s decision was the pivot point. With respect to faculty perspectives, before the decision, universities focused on “student success, finances and organizational structure, and quality assurance” (Kelderman 2023, 39), emphasizing not just admissions but also “faculty hiring [...] and even faculty-promotion” (Whittington 2023, 48). Faculty also expressed critical opinions of DEI efforts. They thought, for instance, that DEI workshops and “mandatory diversity trainings” adopted the same corporate logic they were meant to overcome (Ryan 2023, 43; see also

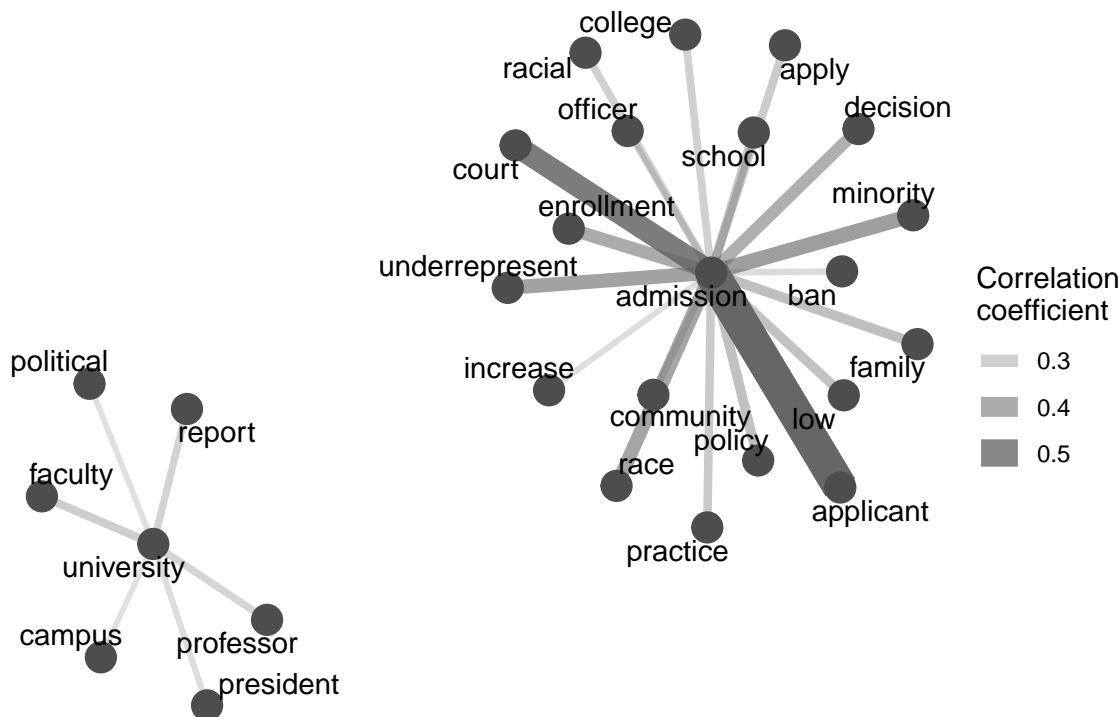


Figure 9: Correlation analysis of the words *university* and *admission* in the *Chronicle of Higher Education*, April 2019–December 2023.

Khalid 2021; Kelderman 2023), an idea echoed in the broader academic literature (Anbrica and Andrews 2025). After the Supreme Court decision, their focus shifted more squarely to student admissions: “The story of the ruling’s aftermath has many levels. It’s about recruitment strategies, admission requirements, and legal compliance” (Hoover 2023, 12).

With respect to students, before the Supreme Court decision, authors emphasized the history of student activism: “Demands for diversity training and other DEI initiatives, such as bias response teams, have been central to student protests against racial injustice since 2015 and have only proliferated in the wake of George Floyd’s murder” (Khalid 2021, 44). (Here and throughout the corpus George Floyd was a constant touchpoint; his name appeared 31 times across 21 of the 118 articles in the corpus.) After the decision, authors focused more on applicants than on students already enrolled; minority applicants in particular expressed anxiety about how their applications would be received and whether they should discuss their experience as members of racialized communities (Hoover 2023).

2024: *Hamis* and *antisemitism*

Congruence

Articles in 2024 were dominated by two events at the end of 2023: the attack launched by Hamas in Israel on October 7, and the Congressional hearings in December about antisemitism at elite universities, especially Harvard and Columbia. To focus on these events, we mapped the words appearing near *Hamis* and *antisemitism* (figure 10).

The second-order interpretants reflected these two events, as well as their interconnection. For instance, *Palestinian*, *Jewish*, and *Israel* belong to both clusters, as do *free* and *speech*, which often appeared as *free*

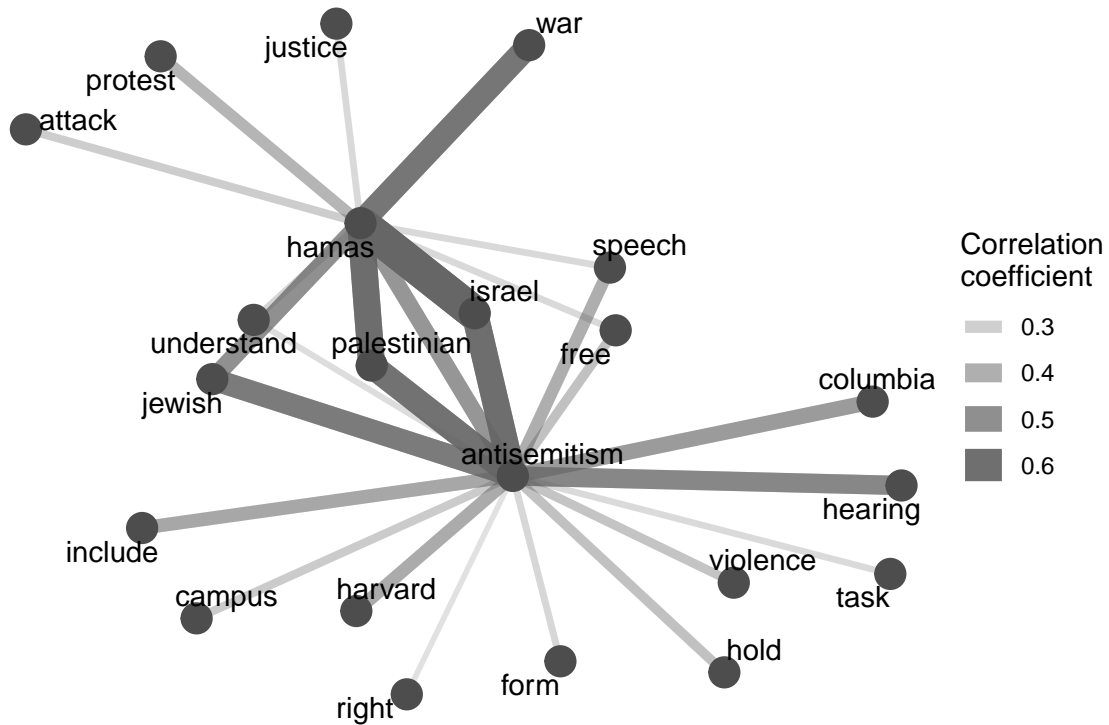


Figure 10: Correlation analysis of the words *Hamas* and *antisemitism* in the *Chronicle of Higher Education*, January–December 2024.

speech (although *free* also frequently appeared before or after *Palestine*). These links are not surprising: the Congressional hearings took place in large part in response to universities’ responses to campus protests related to the conflict between Hamas and Israel. Hence the other words that appear: the names of two universities whose presidents testified before congress, linked to *antisemitism*, and *justice*, *war*, and *protest*, linked to *Hamas*.

Plenitude

Following the same procedure as above, we extracted the articles containing both *Hamas* and *antisemitism*. One notable characteristic of articles published in 2024 was their metadiscourse: to defend their decisions, university leaders were at pains to define what words meant. Thus, they talked more about meaning as such, in contrast to the periods before and after 2024. “In the post-October 7 world,” one author wrote, “many of the fiercest battles in the campus culture wars have taken a strangely Talmudic form: What is antisemitism?” (Rieder 2024) (Other words that people sought to define were *inclusion*, *diversity*, *belonging*, and *civility*: see Mangan 2024.) Campus leaders, other authors wrote, were in a no-win situation:

If they discipline pro-Palestinian protestors over chants many consider antisemitic, they’re accused of trampling free-speech rights. If they defend the right to demonstrate, they’re accused of failing to protect Jewish students from antisemitism. Impartial stances are attacked as weak, sparking debates about whether campus leaders should comment at all. (Hicks and Mangan 2024, 12)

Articles described declining trust and increasing polarization; a rabbi visiting Harvard described universities as “screaming echo-chambers” (quoted in Hicks and Mangan 2024, 17). The governing logic was one of

“vindictive protectiveness,” “hostile solidarity,” and “competing claims to vulnerability,” where speech was marked by a hyperbolic style characterized by metaphors of violence (Gutkin 2024a, 25–6). Len Gutkin, for instance, argued that when “scholars accuse other scholars they disagree with of ‘violence,’ they are not so much making moves in an argumentative game as attempting to call a halt to the game altogether” (2024b, 48).

Articles also described the political opportunism that many perceived behind the Congressional hearings about antisemitism. Universities, Karen Fischer (2024) observed, were “squeezed from both sides”: Republicans were upset with DEI efforts, while Democrats had long been concerned about runaway tuition costs. Republican attacks made the links between the hearings and DEI efforts clear. While the hearings were taking place, Representative Dan Crenshaw, for example, said, “Make no mistake – the DEI bureaucracy is directly responsible for a toxic campus culture that separates everyone into oppressor vs. oppressed” (quoted in Hicks and Mangan 2024, 15).

In this respect, articles from 2024 marked a shift from those that appeared earlier. They focused more on conflict (as observed in the measurements of tone in the sentiment analysis), reflecting a rearrangement of second-order interpretants: whereas before 2024, the initial representamens *diversity*, *DEI*, and *woke* evoked ideas related to admissions, which in turn evoked ideas related to policies addressed by the Supreme Court, over the course of 2024, they evoked ideas related to discrimination in the sociopolitical context defined by the conflict between Israel and Hamas, which in turn evoked ideas related to protests and accusations of antisemitism.

2025: *Title VI* and *Title IX*

Congruence

Thus far, we have observed continuity – the frequent mention of George Floyd, for instance – as well as change. The final period, covering the first five months of 2025, reveals a similar dynamic, shaped by the re-election of President Donald Trump in November 2024. Articles were dominated by discussions of Trump’s policies with respect to Title VI, established in 1964 to ban race-based discrimination, and Title IX, established in 1972 to ban sex-based discrimination. However, these two policies were also touchpoints throughout the corpus: the first articles mentioning Title IX or Title VI appeared in February 2023, the last in May 2025. To identify relevant relationships, we mapped words appearing near *Title VI* and *Title IX* (figure 11).

One notable feature of this map is that the cluster around *Title VI* is larger than the one around *Title IX*, subsuming all but one word. The *Title VI* cluster also contains words that the *Title IX* cluster does not: *affirmative*, *action*, *legal*, *department*, *Trump*, and *race*, suggesting a shift in emphasis away from sex-based discrimination toward race-based discrimination.

Plenitude

One point of continuity from earlier periods is signalled by the words *Jewish* and *Palestinian* in both clusters. The effects of the October 7 attacks and the subsequent campus protests were still being felt. President Biden invoked Title VI protections over the course of 2024 (hence *Biden* in the correlation map), although both Jewish groups and pro-Palestinian advocates found his response inadequate (Bellows 2025). As one article explained, “some colleges now face lawsuits from students on opposing sides of the same events – Muslim students and pro-Palestinian protesters, as well as Jewish students and pro-Israel activists – each claiming the college unfairly favored the other side” (Ford 2025). The response by the Trump administration, starting immediately after Trump’s inauguration in January 2025, was more forceful, marked by the cancellation of millions of dollars of grants at Columbia University and elsewhere, ostensibly for failing to address concerns related to the enforcement of Title VI.

University officials and article authors were interpreting changes to Title VI enforcement through universities’ experience with Title IX more than a decade earlier. (This transposition explains the overlap in clusters

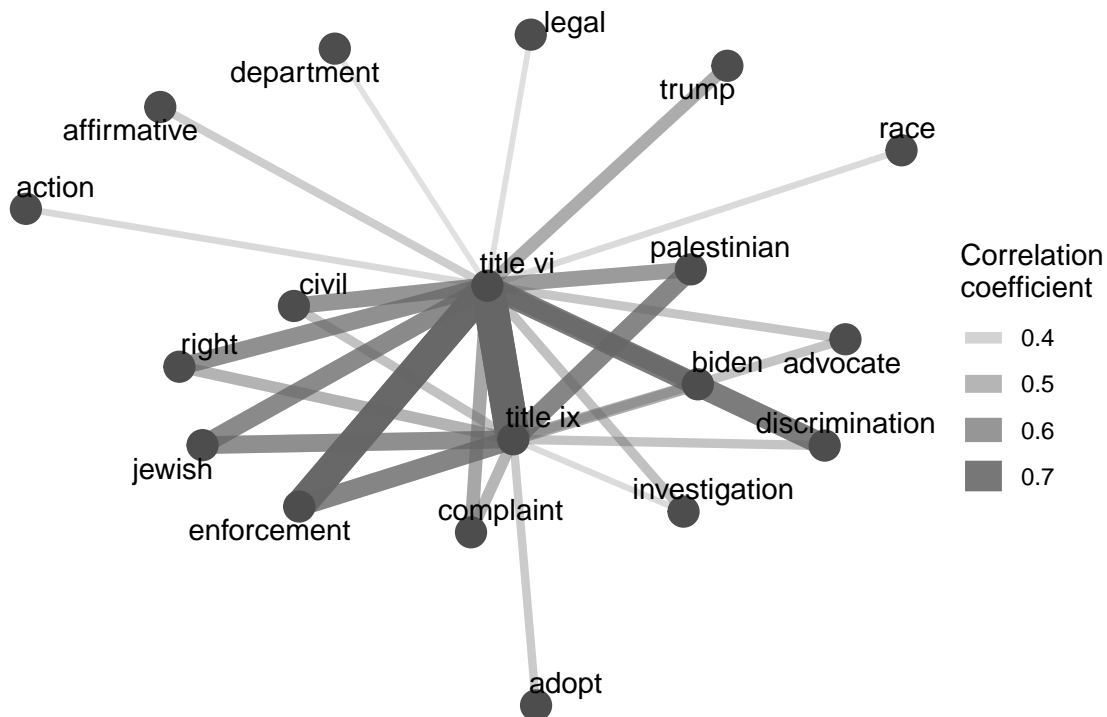


Figure 11: Correlation analysis of the words *Title VI* and *Title IX* in the *Chronicle of Higher Education*, January–May 2025.

in figure 11.) Prior to 2011, Title IX had been “largely regarded as a sports law,” ensuring that women’s teams were supported in ways equivalent to men’s (Bellows 2025). The Obama administration put a new emphasis on victims’ rights, prompting universities to create new offices – and new bureaucracies – to consider discrimination based on sex or gender.

With respect to Title VI, however, Trump went further, according to Richard Thompson Ford (2025), using Title VI as a tool to attack DEI efforts, “turn[ing] antidiscrimination law against itself.” Ford argued that Trump’s policies treated efforts to achieve equity as a zero-sum game: if a minority group received help that the majority group did not, the majority, according to the Trump administration’s interpretation of Title VI, was unjustly penalized.

Interestingly, this critique led Ford to offer a thoroughgoing defence of DEI policies, a useful check on the pessimism characterizing articles about diversity efforts under a Trump presidency (symptoms of which, again, are visible in our earlier sentiment analysis). “It’s no surprise that the anti-college message is winning,” Ford writes: “so far it’s the only message.” Furthermore, “Capitulation to the anti-DEI agenda is unlikely to help colleges avoid” investigations on anti-DEI pretexts, while “[c]ravenly abandoning values that [professors] proudly extolled only a few years ago will ensure that” morale will fall and the “still-considerable prestige of American colleges” will be undermined (Ford 2025). Instead what was called for, he said, was a more robust defence of DEI.

Conclusion: The meanings of diversity

We opened this article by describing how words related to diversity, including *DEI* but especially *woke*, are essentially contested concepts, defined not by an identifiable core meaning, but by “endless disputes” about

come to shape our inquiry itself.

Author contributions

This project is the result of a program intended to introduce undergraduate students to the processes of knowledge creation. The first author is the professor who oversaw the project; the second and third authors are students.

The corresponding author is KC.

Author roles:

- Conceptualization: KC
- Data curation: SD, MLB, KC
- Formal analysis: SD, MLB, KC
- Funding acquisition: none
- Investigation: KC
- Methodology: KC
- Project administration: KC
- Resources: KC
- Software: KC
- Supervision: KC
- Validation: KC
- Visualization: KC
- Writing – original draft: KC
- Writing – review and editing: KC, SD, MLB

References

- Abrica, Elvira Julia, and Ruth Oliver Andrew. 2025. “The Racial Politics of Diversity, Equity, and Inclusion (DEI) Work.” *Journal of Diversity in Higher Education* 18, supplement 1: S259–68. <https://doi.org/10.1037/dhe0000566>.
- Aledavood, Parham. 2024. “Taking the Middle Road: Reflections on Mixed Methodology within the Digital Humanities.” *Digital Studies / Le champ numérique* 14, no. 1: 1–19. <https://doi.org/10.16995/dscn.11069>.
- Asen, Robert. 2024. “Anti-woke Publics.” *Political Communication* 41, no. 6: 1029–34. <https://doi.org/10.1080/10584609.2024.2425782>.
- Bakhtin, Mikhail. 1986. *Speech Genres and Other Late Essays*. Translated by Vern W. McGee. University of Texas Press.
- Bellows, Kate Hidalgo. 2025. “Is Title VI the New Title IX?” *Chronicle of Higher Education*, March 14. Gale Academic OneFile document number A845060936.
- Bock, Tim. 2011. “Improving the Display of Correspondence Analysis Using Moon Plots.” *International Journal of Market Research* 53, no. 3: 307–26. <https://doi.org/10.2501/IJMR-53-3-307-326>.
- Cammaerts, Bart. 2022. “The Abnormalisation of Social Justice: The ‘Anti-woke’ Culture War Discourse in the UK.” *Discourse and Society* 33, no. 6: 730–43. <https://doi.org/10.1177/09579265221095407>.
- Cherry-McDaniel, Monique. 2017. “#Woke: Employing Black Textualities to Create Critically Conscious Classrooms.” *English Journal* 106, no. 4: 41–6. <https://www.jstor.org/stable/26359461>.
- Chronicle of Higher Education*. 2025a. “About Us.” 25 March. <https://www.chronicle.com/page/about-us/>.

- Chronicle of Higher Education*. 2025b. “Broad Reach and Deep Engagement with Every Part of Higher Ed.” 25 March. <https://www.chronicle.com/marketing-solutions>.
- Conway, Kyle. 2023. *How to Read Like You Mean It*. Athabasca University Press. <https://doi.org/10.15215/aupress/9781771993746.01>.
- Dhoest, Alexander, and Steve Paulussen. 2025. “The Mediated Construction of ‘Woke’: Emerging Discourses and Primary Definers in the Flemish Press.” *Journalism* 26, no. 9: 1918–36. <https://doi.org/10.1177/14648849241260942>.
- Fekete, Liz. 2022. “Who Is Behind the ‘War on Woke’: An Interview with Ralph Wilson and Isaac Kamola.” *Race and Class* 64, no. 2: 38–54. <https://doi.org/10.1177/03063968221127591>.
- Field, Kelly. 2023. “A Crusade to End ‘Reverse Discrimination.’” *Chronicle of Higher Education*, July 21: 14–21. Gale Academic OneFile document number A759499611.
- Firth, J.R. 1968. “A Synopsis of Linguistic Theory, 1930–55.” In *Selected Papers of J.R. Firth, 1952–59*, edited by F.R. Palmer, 168–205. Indiana University Press. <https://archive.org/details/selectedpapersof0000firt/page/168/>.
- Fischer, Karin. 2024. “Squeezed from Both Sides.” *Chronicle of Higher Education*, October 4. Gale Academic OneFile document number A824159333.
- Fon, Nguh Nwei Asanga. 2024. “The ‘Woke’ Way or the Highway: American Democracy in the Age of ‘Wokism’ and ‘Cancel Culture.’” *SN Social Sciences* 4: article 90. <https://doi.org/10.1007/s43545-024-00886-w>.
- Ford, Richard Thompson. 2025. “Gutting DEI Won’t Save Us.” *Chronicle of Higher Education*, April 11. Gale Academic OneFile document number A845060818.
- Gallie, W.B. 1956. “Essentially Contested Concepts.” *Proceedings of the Aristotelian Society* 56: 167–98. <https://www.jstor.org/stable/4544562>.
- Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21: 267–97. <https://doi.org/10.1093/pan/mps028>.
- Gutkin, Len. 2024a. “A Decade of Ideological Transformation Come Undone.” *Chronicle of Higher Education*, January 5: 18–27. Gale Academic OneFile document number A780410312.
- Gutkin, Len. 2024b. “The Hyperbolic Style in American Academe.” *Chronicle of Higher Education*, February 16: 44–9. Gale Academic OneFile document number A785993185.
- Harris, Zellig S. 1954. “Distributional Structure.” *Word* 10, nos. 2–3: 146–62. <https://doi.org/10.1080/00437956.1954.11659520>.
- Hicks, Maggie, and Katherine Mangan. 2024. “Can Colleges Protect Jewish Students?” *Chronicle of Higher Education*, January 5: 10–17. Gale Academic OneFile document number A780410310.
- Hogan, Kelly A., and Viji Sathy. 2023. “How Can ‘Inclusion’ Be a Bad Word?” *Chronicle of Higher Education*, April 14: 52–3. Gale Academic OneFile document number A757106730.
- Hoover, Eric. 2023. “A Time to Tear Down, a Time to Build Up.” *Chronicle of Higher Education*, October 13: 10–25. Gale Academic OneFile document number A772675914.
- Hu, Minqing, and Bing Liu. 2004. “Mining and Summarizing Customer Reviews.” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD-2004)*: 168–77. <https://doi.org/10.1145/1014052.1014073>.
- Iyer, Aarti. 2022. “Understanding Advantaged Groups’ Opposition to Diversity, Equity, and Inclusion (DEI) Policies: The Role of Perceived Threat.” *Social and Personality Psychology Compass* 16: article e12666. <https://doi.org/10.1111/spc3.12666>.

- Kelderman, Eric. 2023. “The New Accountability.” *Chronicle of Higher Education*, April 14: 36–43. Gale Academic OneFile document number A757106727.
- Khalid, Amna. 2021. “Demands for Diversity Lead to Corporatization.” *Chronicle of Higher Education*, May 14: 44–5. Gale Academic OneFile document number A663389766.
- Leetaru, Kalev Hannes. 2012. *Data Mining Methods for the Content Analyst: An Introduction to the Computational Analysis of Content*. Routledge. <https://doi.org/10.4324/9780203149386>.
- Mangan, Katherine. 2023. “Doubling Down on DEI.” *Chronicle of Higher Education*, July 7: 24–31. Gale Academic OneFile document number A757798350.
- Mangan, Katherine. 2024. “The End of Inclusion?” *Chronicle of Higher Education*, March 1: 40–5. Gale Academic OneFile document number A786212633.
- Markham, Annette N. 2013. “Undermining ‘Data’: A Critical Examination of a Core Term in Scientific Inquiry.” *First Monday* 18, no. 10. <https://doi.org/10.5210/fm.v18i10.4868>.
- McCurdy, Patrick, Kaitlin Clarke, and Bart Cammaerts. 2025. “From Social Awareness to Authoritarian Other: The Conservative Weaponization of Woke in Canadian Parliamentary Discourse.” *Journal of Language and Politics* 24, no. 6: 910–33. <https://doi.org/10.1075/jlp.24126.mcc>.
- McWhorter, John. 2021. *Woke Racism: How a New Religion Betrayed Black America*. Penguin. ProQuest Ebook Central.
- Mogilski, Justin, Lee Jussim, Anne Wilson, and Bryan Love. 2025. “Defining Diversity, Equity, and Inclusion (DEI) by the Scientific (De)merits of Its Programming.” *Theory and Society*. <https://doi.org/10.1007/s11186-025-09646-y>.
- Mohammad, Saif, and Peter Turney. 2013. “Crowdsourcing a Word-Emotion Association Lexicon.” *Computational Intelligence* 29, no. 3: 436–65. <https://doi.org/10.1111/j.1467-8640.2012.00460.x>.
- Moretti, Franco. 2000. “Conjectures on World Literature.” *New Left Review* 1: 54–68.
- Nielsen, Finn Årup. 2011. “A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs.” *Proceedings of the ESWC2011 Workshop on “Making Sense of Microposts”: Big Things Come in Small Packages*: 93–8. <http://arxiv.org/abs/1103.2903v1>.
- O’Sullivan, David, and David J. Unwin. 2010. *Geographic Information Analysis*, 2nd ed. Wiley. <https://doi.org/10.1002/9780470549094>.
- Peirce, C.S. 1940. “Logic as Semiotic: The Theory of Signs.” In *The Philosophy of Peirce: Selected Writings*, edited by Justus Buchler, 98–119. Routledge. <https://doi.org/10.4324/9781315822723>.
- Piper, Andrew. 2015. “Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel.” *New Literary History* 46, no. 1: 63–98. <https://doi.org/10.1353/nlh.2015.0008>.
- R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Reisch, Michael, and Jayshree S. Jani. 2025. “Deconstructing DEI: Unmasking Its Complexities, Contradictions, and Challenges.” *Journal of Teaching in Social Work* 45, no. 2: 250–75. <https://doi.org/10.1080/08841233.2025.2469550>.
- Ricoeur, Paul. 1981. *Hermeneutics and the Human Sciences: Essays on Language, Action and Interpretation*. Translated and edited by John B. Thompson. Cambridge University Press. <https://doi.org/10.1017/CBO9781316534984>.
- Rieder, Jonathan. 2024. “Activist Professors at Columbia and Barnard Are Botching Free Speech.” *Chronicle of Higher Education*, May 10. Gale Academic OneFile document number A795480550.
- Rinker, Tyler W. 2018. *textstem: Tools for Stemming and Lemmatizing Text*. V. 0.1.4. <http://github.com/trinker/textstem>.

- Roberts, Jasmine. 2021. “‘Woke Culture’: The Societal and Political Implications of Black Lives Matter Digital Activism.” In *Democracy in the Disinformation Age: Influence and Activism in American Politics*, edited by Regina Luttrell, Lu Xiao, and Jon Glass, 37–57. Routledge. <https://doi.org/10.4324/9781003008828>.
- Ryan, James E. 2023. “DEI: The Case for Common Ground.” *Chronicle of Higher Education*, June 9: 42–3. Gale Academic OneFile document number A757639958.
- Salton, Gerard, and Christopher Buckley. 1988. “Term-Weighting Approaches in Automatic Text Retrieval.” *Information Processing and Management* 24, no. 5: 513–23. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
- Scholz, Ronny. 2019. “Lexicometry: A Quantifying Heuristic for Social Scientists in Discourse Studies.” In *Quantifying Approaches to Discourse for Social Scientists*, edited by Ronny Scholz, 123–53. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-97370-8_5.
- Silge, Julia, and David Robinson. 2016. “tidytext: Text Mining and Analysis Using Tidy Data Principles in R.” *Journal of Open Source Software* 3, no. 1. <https://doi.org/10.21105/joss.00037>.
- Silge, Julia, and David Robinson. 2017. *Text Mining with R: A Tidy Approach*. O’Reilly. <https://www.tidytextmining.com>.
- Vološinov, V.N. 1986. *Marrxism and the Philosophy of Language*. Translated by Ladislav Matejka and I.R. Titunik. Harvard University Press.
- Weaver, Warren. 1949. “The Mathematics of Communication.” *Scientific American* 181, no. 1: 11–15. <https://www.jstor.org/stable/10.2307/24967225>.
- Whittington, Keith E. 2023. “A Governor’s Plot Against Higher Ed.” *Chronicle of Higher Education*, March 17: 48–9. Gale Academic OneFile document number A757639920.
- Wiedmaier, Benjamin. 2018. “Phi Coefficient.” In *SAGE Encyclopedia of Communication Research Methods*, 1231–32. Sage. <https://doi.org/10.4135/9781483381411.n427>.