*Corresponding Author's Email: huyducnguyen.khxhvnv@gmail.com
Proceedings of the International Conference on Research in Social Sciences
Vol. 3, Issue. 1, 2026, pp. 1-14
DOI: https://doi.org/10.33422/rssconf.v3i1.2093
Copyright © 2026 Author(s)
ISSN: 3030-0452 online



Diamond Scientific Publishing

The Polysemy of "Bias" in Contemporary English: A Corpus-Based Disambiguation Analysis Across Social, Statistical, Juridical, Technical, and Everyday Registers

Duc Huy Nguyen[1*] and Nguyen Khang Nguyen[2]

[1] Van Hien University, Vietnam
[2] Ho Chi Minh City University of Foreign Languages and Information Technology (HUFLIT), Vietnam

Abstract

In contemporary English, "bias" demonstrates several meanings that change with each register, as the same form of "bias" indicates different meanings. It is a social prejudice, a procedural ground to disqualify a judge, or a systematic error in research. Therefore, this study explores the distribution of these senses in political, legal, and daily English contexts, as well as what that distribution implies for public comprehension. Combining etymological reconstruction with corpus analysis, it operationalized the eight Cambridge Dictionary senses as a coding inventory and assembled eight sub-corpora from the english-corpora.org family, supplemented with United States court opinions and a daily-usage benchmark. Collocates within an L5 to R5 window were ranked by Log-Likelihood, with Mutual Information as a supplementary measure, and sense membership was tested through a chi-square test of independence with Cramér's V and Monte Carlo confirmation. The seven sense profiles proved statistically separable, $\chi^2(305, N = 6{,}922) = 13{,}559.35$, $p < .001$, Cramér's $V = .62$, the sharpest separation falling in the statistical sense and the weakest between the prejudice and political-legal senses. In everyday usage the preference reading predominated while the physical senses receded, concentrating comprehension risk where the civic stakes of "bias" are highest.

Keywords: bias, collocational analysis, corpus linguistics, lexicography, polysemy

1 Introduction

1.1 Statement of Problem

The English lexeme "bias" presents a complex case of polysemic expansion across markedly different domains. The word entered English in the 1520s as a noun for an oblique or diagonal

line, migrated into figurative legal usage by the 1570s as a one-sided tendency of the mind, and by the early twentieth century had acquired a technical scientific sense of inaccuracy in research results.

The etymological spread of "bias" matters in political and legal discourse because a single form carries sharply different meanings that depend on the register in which readers meet it. In legal contexts "bias" denotes a failure of judicial impartiality, addressed through recusal motions and disqualification. Rädler (1997) affirms that "in English law and the law of many Commonwealth countries the rule against bias is considered to be a principle of 'natural justice,'" applying to each member of the bench and of the jury. Mills (1999) defines judicial bias as "the opposite of judicial impartiality," involving prejudiced "feelings or spirit" toward claimants that judges are obligated to expunge from their reasoning (p. 12).

In political and media discourse the word denotes ideological slant, partisan framing, or the systematic favouring of one position over another. Chen et al. (2020) show that analysis of article excerpts reveals "insightful bias patterns at different levels of text granularity, from single words to the whole article discourse" (p. 149). The Australian Law Reform Commission (2021) notes that cognitive shortcuts, though "generally useful and allow us to function," also "bias decision-making and can lead to error." In research methodology the word denotes inaccuracy arising when a factor is left uncontrolled, a sense distinct from random error: as Weisberg (2010) emphasizes, "Unlike error related to random variability, bias cannot be assessed without external knowledge of the world."

The communicative consequences of this polysemy are not trivial. In discussions of electoral fairness, "partisan bias" is often taken as a single concept when it in fact layers several. Ramsay (2025) observes that "the colloquial phrase 'partisan bias' encompasses multiple distinct conceptions of bias, including partisan advantage, packing & cracking, and partisan symmetry" (p. 1), and characterizes the term as multiordinal, an abstract umbrella covering at least three distinct types of bias.

Curley et al. (2022) document that "in ambiguous environments, jurors, who are not legal experts, may be overwhelmed by the legal terminology being directed at them," so that this "cognitively taxing environment may cause jurors to have a low elaboration likelihood and thus utilise bias when reaching their verdict." Rodrigo-Ginés et al. (2024) add that bias surfaces through "lexical features pertaining to word choice and usage, morphosyntactic features related to grammar and sentence structure, and semantic features dealing with meaning," and warn that media bias "pose a threat to the proper functioning of democratic processes."

Despite the civic stakes attached to this semantic micro-choice, the linguistic foundations of how "bias" functions across registers in present-day English remain underexplored in publicly accessible corpus-based scholarship. Existing research has examined "bias" as a psychological construct, a measurable feature of media output, and a procedural ground for judicial recusal, but its distribution across political and legal registers has not been documented as a register-sensitive corpus profile.

This study addresses that gap by asking how the polysemy of "bias" is distributed across registers in present-day English and what that distribution means for societal comprehension of political and legal communication. When readers meet the word in political reporting or legal documents, do they process it as a personal preference, an ideological slant, a disqualifying legal predisposition, or a statistical deviation? The study translates its findings into recommendations for journalism, legal drafting, and public legal education.

1.2 Research Questions

1. How has the polysemic structure of "bias" evolved from its etymological origin into its present-day senses in English?
2. How does the distribution of senses of "bias" in political and legal discourse differ from the distribution found in general everyday English in present-day usage?
3. Which collocational patterns to the left and to the right of the node "bias" most reliably disambiguate the eight documented senses in authentic running text?
4. How can the observed sense distributions and collocational patterns of "bias" inform public comprehension of political journalism and legal communication?

This review sets out the theoretical and methodological basis for a register-sensitive account of the polysemy of "bias," drawing together lexicographic, cognitive, legal, political, and methodological scholarship with the corpus-linguistic tradition supporting the analysis, on the principle that a distributional account of polysemy is most secure when each sense has already been characterized on independent grounds (Nguyen et al., 2016).

1.3 Lexicographic and Etymological Accounts of Bias

The historical pathway of "bias" is documented most fully in the Online Etymology Dictionary. In the 1520s "bias" appears as a noun for an oblique or diagonal line, borrowed from Middle French "biais" and traced through Old Provençal to Greek "epikarsios," meaning athwart or crosswise, with the root "sker," to cut (Harper, n.d.; Aronson, 2018). Its earliest uses referred to a diagonal cut in fabric and to the asymmetric construction of bowls, as Aronson (2018) notes: "Both the jack and the bowls are asymmetrically shaped, and this asymmetry, known as bias, distorts the direction in which they travel - in a curve, not a straight line." Harper records its migration into bowls in the 1560s, into figurative legal and rhetorical usage in the 1570s, into a transitive verbal sense in the 1620s, and into the sense "depict unfairly" in 1872.

The Cambridge Dictionary records eight present-day senses, from unfairly influenced judgement to the statistical, bowls, fabric, and verbal senses (Cambridge University Press, n.d.). Oxford Learner's Dictionaries (Oxford University Press, n.d.) and Merriam-Webster (n.d.) confirm these, the latter also recording a specialised electronics sense, furnishing the independent inventory against which authentic instances are classified.

1.4 The Social, Political, and Legal Senses of Bias

Jurisprudence treats the legal sense as a failure of judicial impartiality capable of disqualifying a decision-maker. Rädler (1997) locates the rule against bias within natural justice and extends it to every member of the bench and jury, Mills (1999) defines it as the inverse of impartiality, and Geyh (2013) and Nugent (1994) reach similar conclusions. The cognitive, media, and methodological literatures cited above treat "bias" instead as a systematic tendency shaped by prior attitudes, detectable in text and controllable by design, which makes it a social and political problem rather than a stylistic one.

1.5 Lexical Ambiguity and Systematic Error

Legal linguistics treats lexical indeterminacy as a practical problem. Solan (1993) argues that judicial interpretation often turns on unstated ordinary-language intuitions, and Tiersma (1999) shows that legal English contains "legal homonyms," words resembling ordinary vocabulary but carrying distinct legal meanings. Schane (2006) separates lexical ambiguity, referential ambiguity, and vagueness. This bears directly on "bias," whose procedural legal sense is technically defined yet shares its surface form with a socially loaded everyday sense, a configuration in which a reader may recover the wrong sense unawares. The statistical sense

rests on the methodological literature on systematic error: Weisberg (2010) separates bias from random variability, and Hernán and Robins (2020) distinguish selection, measurement, and confounding bias from chance variation, treating "bias" as a non-random distortion of measurement, estimation, sampling, or inference that study design attempts to control.

1.6 Corpus-Linguistic Foundations for the Study of Polysemy

Sinclair (1991) establishes corpus, concordance, and collocation as the empirical basis for describing patterned co-occurrence, meaning being distributed across recurrent lexical environments rather than isolated forms; Hunston (2002) sets out the conventions for assessing collocational attraction, including Mutual Information and Log-Likelihood; and McEnery and Hardie (2012) articulate reproducibility standards, with english-corpora.org (Davies, 2008) and AntConc (Anthony, 2024) providing the apparatus. Computational approaches differ: Navigli (2009) surveys word sense disambiguation as assigning a label from a fixed inventory while Camacho-Collados and Pilehvar (2018) show that unsupervised representations can induce distinctions without one. The present study takes the lexicographic route, trading coverage for auditability of every classification decision.

1.7 Synthesis and the Contribution of the Present Study

Three points follow. The senses of "bias" are independently attested in both the lexicographic record and the domain literatures; the multiordinal character of the word motivates a register-stratified analysis and predicts overlap between the prejudice and political-legal senses; and the corpus-linguistic tradition (Sinclair, 1991; Hunston, 2002; McEnery & Hardie, 2012) supplies the apparatus.

2 Methodology

2.1 Research Design

The study follows a quantitative design combining two complementary analyses: a lexicographic and etymological reconstruction of the sense inventory of "bias," and a corpus-based collocational analysis across eight register-specific sub-corpora, so that each corpus sense is interpreted against an independently established benchmark.

2.2 Data Collection

In phase 1, the etymology and lexicography of "bias" are reconstructed from the Online Etymology Dictionary and the Cambridge, Oxford, and Merriam-Webster dictionaries, which supply the independent sense inventory. In phase 2, corpus data are drawn from the english-corpora.org platform (Davies, 2008), chiefly the News on the Web Corpus and the Corpus of Contemporary American English, with publicly available United States federal court opinions supplying procedural legal usage and the global web-based corpora supplying an everyday baseline. Eight sub-corpora were compiled, one for each target sense plus a register-free daily-usage set, and indexed in AntConc 4.3.1 (Anthony, 2024). For each, concordance lines were extracted within an L5 to R5 window, function words were removed by stoplist, and content collocates were ranked by the four-term Log-Likelihood, with Mutual Information as a complementary index. Only collocates reaching the critical value of 3.84, corresponding to $p < .05$, were retained; where a sub-corpus was too small for any collocate to reach it, content collocates were tabulated from the full window with a custom R routine applying the same stoplist and are reported as absolute counts. All inferential statistics were computed in R 4.5.1 (R Core Team, 2025).

3 Data Analysis and Discussion

3.1 Data Analysis

3.1.1 Lexicographic Analysis of the Concept of "Bias"

Table 1. Diachronic Synthesis of "Bias" Across Two Reference Works

Period	Source	Core sense	Extensions
13th century	Etymology (Old French and Old Provençal biais, via Greek epikarsios)	Sideways, askance, against the grain	Athwart, crosswise, at an angle; cognate forms in Old Catalan and Sardinian
1520s	Etymology (English, from French biais)	Oblique or diagonal line	An expedient or means, in figurative use
1560s–1570s	Etymology (English)	A technical term in the game of bowls: a ball weighted on one side so that it curves obliquely	A one-sided tendency of the mind, especially in legal use
1620s–1680s	Etymology (English)	As a verb, to give a bias to, or to cause to incline to one side	A slant or deviation, correlated with the word skew
Modern	Cambridge	An unfair support of, or opposition to, a person or thing that arises when personal opinion overrides fair judgement	A preference for a subject or thing; a statistical skew in data; the shape of a bowls ball; a diagonal across woven fabric; a verbal sense; and a political and legal partiality

Note. Historical rows follow the Online Etymology Dictionary (n.d.); the modern row follows the Cambridge Dictionary (n.d.).

The three largest sets, Prejudice, Political and Legal, and Statistical, support inferential statistics, while the Preference, Bowls, Fabric, and Verb sets are small enough to require descriptive treatment as the sense profiles below show.

3.1.2 Corpus-based Analysis

The eight sub-corpora described in Section 2.2 are analysed here with the association measures and thresholds defined there. Table 2 reports their composition

Table 2. Sub-Corpus Profile of the Eight Bias Datasets

Sub-corpus	Source file	Tokens	Types	KWIC instances
Sense 1 (Prejudice)	BIAS PREJUDICE AND DISCRIMINATION.docx	139,352	15,453	5,511
Sense 2 (Preference)	BIAS_PREFERENCE.docx	2,176	729	114
Sense 3 (Statistical)	BIAS_STATISTICAL_DATA.docx	49,739	5,804	2,477
Sense 4 (Bowls)	BIAS_IN_BOWL_GAME.docx	40	33	3
Sense 5 (Fabric)	BIAS_FABRIC.docx	1,904	571	100
Sense 6 (Verb)	BIAS_VERB.docx	329	189	15
Sense 7 (Political/Legal)	BIAS_POLITICAL_LEGAL.docx	61,928	7,774	2,425
Daily Usage	BIAS_DAILY_USAGE.docx	515	240	28
Total	—	**255,907**	—	**13,189**

Note. Tokens and types are AntConc indexer counts under the default token definition; KWIC instances are concordance lines retrieved with "bias" as the node. Sub-corpus size reflects the frequency of each sense in available text rather than a sampling target.

3.1.3 Sense 1: Prejudice

The Prejudice sub-corpus contains 139,352 tokens and produced 5,511 KWIC instances of "bias," the largest sense-specific dataset, its L5 to R5 environment dominated by collocates marking social categories, ideological positions, and unfair judgement. Table 3 reports the fifteen strongest, headed by confirmation, against, racial, liberal, and gender. The strongest Mutual Information bonds are confirmation (MI = 1.50), attentional (MI = 1.37), implicit (MI

= 1.36), and racial (MI = 1.22), all below Hunston's (2002) benchmark of 3.0, as expected for a high-frequency node, so Log-Likelihood is the principal criterion.

Table 3. Content Collocates of Bias in the Prejudice Sub-Corpus

Collocate	Freq.	Freq. L	Freq. R	Log-Likelihood	MI
confirmation	138	130	8	116.19	1.50
against	275	30	245	110.42	0.99
racial	179	164	15	103.55	1.22
liberal	163	154	9	78.28	1.10
gender	163	151	12	75.44	1.07
media	247	177	70	67.58	0.80
toward	140	5	135	64.17	1.07
anti	113	107	6	55.00	1.10
selection	87	78	9	42.72	1.11
implicit	56	50	6	39.72	1.36
attentional	46	45	1	32.86	1.37
towards	67	1	66	31.50	1.08
political	111	85	26	26.67	0.75
risk	67	58	9	26.24	0.98
potential	67	58	9	25.43	0.96

Note. Measures as defined in Section 2.2.

The concordance lines confirm the prejudice sense as dominant. Confirmation bias instantiates the first Cambridge sense, while racial, gender, liberal, and implicit bias broaden it to social-category and ideological contexts. The frequency columns separate two groups: social-category and cognitive modifiers cluster to the left of the node, while directional prepositions and "media" fall predominantly to its right.

3.1.4 Sense 2: Preference

The Preference sub-corpus contains 2,176 tokens and produced 114 KWIC instances of "bias," its environment dominated by an indefinite-NP construction rather than social-category premodifiers. Because AntConc returned only two collocates reaching the threshold, the indefinite article a (Log-Likelihood = 16.21) and the self-collocate "bias" (26.02), the content collocates in Table 4 were tabulated from the full L5 to R5 window using a custom R routine applying the same stop list and lemma-handling as for the larger sub-corpora.

Table 4. Content Collocates of Bias in the Preference Sub-Corpus

Collocate	Freq.	Freq. L	Freq. R	Log-Likelihood	MI
confirmation	138	130	8	116.19	1.50
against	275	30	245	110.42	0.99
racial	179	164	15	103.55	1.22
liberal	163	154	9	78.28	1.10
gender	163	151	12	75.44	1.07
media	247	177	70	67.58	0.80
toward	140	5	135	64.17	1.07
anti	113	107	6	55.00	1.10
selection	87	78	9	42.72	1.11
implicit	56	50	6	39.72	1.36
attentional	46	45	1	32.86	1.37
towards	67	1	66	31.50	1.08
political	111	85	26	26.67	0.75
risk	67	58	9	26.24	0.98
potential	67	58	9	25.43	0.96

Note. Measures as defined in Section 2.2, except that no association statistics could be computed; the table reports absolute counts from the R routine. Toobin refers to the legal commentator Jeffrey Toobin.

Three patterns mark Sense 2. The first is the financial compound "home bias" with 22 occurrences, denoting investors favouring domestic equities. The second is confessional, marked by the left collocates admit, personal, and strong with a nearby first-person I. The third is prepositional targeting, as in "bias toward" and "bias in favor of," presenting "bias" as neutral

preference rather than prejudice. This sense is identified by construction rather than by any single collocate.

Sense 3: Statistical

The Statistical sub-corpus contains 49,739 tokens and produced 2,477 KWIC instances of "bias," the third-largest dataset, its environment dominated by collocates from quantitative methodology, epidemiology, and clinical-research design. Table 5 reports the fifteen strongest, headed by selection (Log-Likelihood = 235.41), detection, reporting, publication, and attrition, with compounds such as "selection bias" and "publication bias" recurring as lexicalized terms.

Table 5. Content Collocates of Bias in the Statistical Sub-Corpus

Collocate	Freq.	Freq. L	Freq. R
home	22	17	5
strong	11	7	4
toward	11	0	11
field	10	7	3
towards	10	0	10
against	8	1	7
everyone	8	4	4
favor	8	2	6
admit	7	7	0
instances	6	3	3
personal	6	5	1
someone	6	2	4
expect	5	2	3
Toobin	5	4	1
favor	4	1	3

Note. Measures as defined in Section 2.2. The collocates sec, iso, and mm come from engineering and metrology documents where "bias" denotes measurement error.

Cluster analysis confirms the methodological specialization. Verbal collocates distinguish this sense from all others: reduce (frequency 206), eliminate (131), and minimize occur predominantly to the left of "bias," signalling an imperative to control the phenomenon rather than attribute it, a frame absent from the prejudice and political-legal sub-corpora. This sense also returns the highest Mutual Information values in the study, 2.56 for detection and 2.29 for attrition.

3.1.5 Sense 4 to 6: Bowls, Fabric, and Verb

These three senses are marginal in present-day English and are treated together. The Bowls sub-corpus is the smallest, with 40 tokens and 3 KWIC instances. The Fabric sub-corpus contains 1,904 tokens and 100 KWIC instances centred on "bias binding" (frequency 43) and "bias tape" (26), with the frame cut on the "bias." The Verb sub-corpus contains 329 tokens and 15 KWIC instances of "bias" as a transitive verb; with no collocate reaching the threshold it is identified syntactically, by a preceding modal or causative and a direct-object NP. Table 6 reports the collocates for all three.

Table 6. Content Collocates of Bias in the Bowls, Fabric, and Verb Sub-Corpora

Collocate	Freq.	Freq. L	Freq. R	Log-Likelihood	MI
selection	311	254	57	235.41	1.40
detection	85	45	40	180.30	2.56
reporting	129	69	60	169.97	1.93
publication	204	186	18	142.70	1.34
attrition	78	39	39	137.44	2.29
performance	83	50	33	102.37	1.86
reduce	206	202	4	82.19	0.98
attentional	118	101	17	81.89	1.33
cognitive	59	38	21	77.96	1.94
recall	96	74	22	69.09	1.36
sec	40	19	21	59.79	2.08
iso	40	26	14	59.79	2.08
eliminate	131	128	3	52.74	0.98
mm	41	14	27	52.72	1.90
exposure	68	41	27	44.18	1.29

Note. Measures as defined in Section 2.2, except that no association statistics could be computed for Bowls or Verb; counts come from the R routine across the full window. In the Fabric set, ball reflects metalinguistic comparison with the bowls sense rather than polysemic overlap; instances, corpora, and verb are excluded as header artefacts.

None of the three is identified by a productive collocational profile: Bowls and Fabric survive only inside fixed compounds and the verbal use is recognized by syntactic frame, so all three depend on register restriction rather than semantic contrast.

3.1.6 Sense 7: Political and Legal

The Political and Legal sub-corpus contains 61,928 tokens and produced 2,425 KWIC instances of "bias," the second-largest dataset, its environment dominated by collocates from judicial procedure, electoral commentary, and political journalism. Table 7 reports the thirteen strongest, headed by prejudice, racial, actual, against, and gender. The Log-Likelihood for prejudice (106.47) is notable because it co-occurs overwhelmingly to the right of "bias" (164 of 183 occurrences), signaling the coordination of "bias" or "prejudice" pervasive in legal opinions.

The concordance lines confirm "bias" as a procedurally consequential attitude of a decision-maker that may disqualify them or invalidate a decision, as in "actual bias on the part of the judge." The premodifiers actual, implied, possible, potential, and pro mark the technical legal reading, absent or marginal in the Prejudice sub-corpus. This sense mirrors the statistical one: where Sense 3 treats bias as something to be eliminated, Sense 7 treats it as something to be established and proved.

Table 7. Content Collocates of Bias in the Political and Legal Sub-Corpus

Sub-corpus	Collocate	Freq.	Freq. L	Freq. R
Bowls	bowl	3	1	2
	ball	1	0	1
	game	1	0	1
	half	1	0	1
	huge	1	1	0
Fabric	binding	43	3	40
	tape	26	3	23
	fabric	15	10	5
	cut	10	7	3
	ball	8	2	6
	phrases	8	3	5
	finished	7	3	4
	stretch	7	3	4
	sewing	6	4	2
	cm	5	5	0
	hot	5	5	0
	making	5	5	0
	pennants	5	5	0
	pink	5	5	0
Verb	cause	6	4	2
	outcome	3	0	3
	case	2	0	2
	challenge	2	2	0
	maybe	2	0	2
	population	2	1	1
	ability	1	1	0
	according	1	0	1
	against	1	0	1
	besides	1	0	1
	biasing	1	0	1

Note. Measures as defined in Section 2.2. The negative MI for "the court" reflects structural rather than lexical attraction.

3.1.7 Sense Distribution in Daily Usage

The Daily Usage sub-corpus contains 515 tokens and 28 KWIC instances of "bias." It is register-free, built from online forums, fandom commentary, sports writing, and opinion, and measures the proportional distribution of the seven senses in unfiltered usage. The 28 lines were classified by a cue-based procedure in R using the sense-specific lexis established above. Each line was scored by cue count in an L10 to R10 window and assigned to the highest-scoring sense, marked Ambiguous if two or more tied and Unresolved if none appeared. Table 8 reports the distribution.

Table 8. Sense Distribution of Bias in the Daily Usage Sub-Corpus

Collocate	Freq.	Freq. L	Freq. R	Log-Likelihood	MI
prejudice	183	19	164	106.47	1.22
racial	174	171	3	101.34	1.22
actual	152	142	10	85.03	1.19
against	183	23	160	65.86	0.93
gender	72	66	6	36.61	1.13
court	92	33	59	36.35	−0.81
implied	55	42	13	35.89	1.30
appearance	36	26	10	29.37	1.48
possible	56	49	7	28.79	1.14
probability	38	35	3	24.04	1.28
selection	44	20	24	22.69	1.14
potential	51	49	2	22.62	1.05
pro	28	27	1	19.55	1.35

Note. Sense assignment used automated cue-based classification at L10 to R10, in R 4.5.1. Percentages are computed within the full sample of 28 instances, not the 11 resolved cases.

Of the 11 resolved lines, Preference accounts for 7 (63.6%), Prejudice for 2 (18.2%), Statistical for 1 (9.1%), and Verb for 1 (9.1%); Bowls, Fabric, and Political-Legal have none. A chi-square goodness-of-fit test of the four senses with at least one resolved instance was significant, $\chi^2(3, N = 11) = 9.00$, $p = .029$. Everyday "bias" is now primarily the preference reading.

3.1.8 Chi-Square Test of Sense by Collocate Independence

To test whether the collocational profiles of the seven senses are distinguishable, a Pearson chi-square test of independence was run in R on a table of senses by diagnostic collocates, retaining only collocates with at least five occurrences. The Bowls sub-corpus contained no retained cue, so its all-zero row was excluded before computing Cramér's V (Ben-Shachar et al., 2020) and the test covers six of the seven senses. The final matrix comprised six rows and 62 columns (total 6,922). The result was highly significant, $\chi^2(305, N = 6,922) = 13,559.35$, $p < .001$, Cramér's V = .62, and a Monte Carlo rerun at 10,000 replicates confirmed it at $p < .0001$. Table 9 summarizes both tests.

Table 9. Summary of Inferential Statistical Tests

Test	χ^2	df	p	Cramér's V	N
Chi-square independence (classical)	13,559.35	305	< .001	.62	6,922
Chi-square independence (Monte Carlo, B = 10,000)	13,559.35	—	≈ .0001	.62	6,922
Chi-square goodness-of-fit (Daily Usage)	9.00	3	.029	—	11

Sense category	Frequency	Percentage
Ambiguous (tied score)	12	42.9%
Sense 2: Preference	7	25.0%
Unresolved (no cue in window)	5	17.9%
Sense 1: Prejudice	2	7.1%
Sense 3: Statistical	1	3.6%
Sense 6: Verb	1	3.6%
Sense 4: Bowls	0	0.0%
Sense 5: Fabric	0	0.0%
Sense 7: Political/Legal	0	0.0%
Total	**28**	**100.0%**

Note. Both tests were computed in R 4.5.1 (R Core Team, 2025); Cramér's V from the effectsize package (Ben-Shachar et al., 2020), with Monte Carlo simulation at B = 10,000.

The Cramér's V of .62 is large under Cohen's (1988) benchmarks, and its magnitude reflects the partially overlapping profiles of the Prejudice and Political-Legal sub-corpora, which share social-category premodifiers and directional prepositions. The two civic senses nonetheless stay apart through distinctive cues: actual, implied, possible, and potential mark Sense 7 while confirmation, unconscious, implicit, and attentional mark Sense 1.

3.2 Discussion

Answer to Research Questions

Question 1. "Bias" has evolved from a concrete physical meaning into a layered array of abstract, attitudinal, and technical meanings, as set out in Table 1. The corpus shows the

physical sense largely lost, bowls yielding three instances and fabric surviving only in garment-making.

Question 2. The distribution is sharply register-conditioned. The procedural sense produced 2,425 concordance instances yet no resolved case among the 28 daily-usage instances, as was also true of bowls and fabric. Everyday usage is dominated by preference (63.6% of resolved instances), then prejudice (18.2%) and the statistical and verbal senses (9.1% each), $\chi^2(3, N = 11) = 9.00$, $p = .029$.

Question 3. Each sense carries a distinct collocational signature, and these proved separable in Section 3.1 (Cramér's V = .62). The weakest separation falls along the civic axis, where the prejudice and political-legal senses share racial, gender, and political, with procedural and cognitive cues keeping them apart.

Question 4. The statistical sense carries the lowest comprehension risk. The greatest risk lies where the prejudice and political-legal senses overlap and where everyday readers, defaulting to preference, are least able to recover the procedural meaning. Journalists and legal drafters should therefore supply the disambiguating cues identified here, and public legal education should teach the procedural sense explicitly.

4 Implications and Conclusion

4.1 Implications

Theoretically, tracing "bias" from its sixteenth-century geometrical origin through its figurative, attitudinal, and technical senses shows that trajectory to be detectable in the present corpus distribution. The study extends polysemy analysis to the register-sensitive distribution of a high-frequency civic item, supports the view that meaning resides in recurrent lexical environments rather than in the form (Sinclair, 1991; Hunston, 2002), and supports Ramsay's (2025) multiordinality thesis through the weak separation along the civic axis. Practically, because the legal and political sense is almost absent from everyday communication, journalists cannot assume readers recognise "media bias," where the everyday meaning is preference. For legal drafting the implication is more serious, since the sense used in disqualifying procedure is largely foreign to lay readers such as jurors. The cues identified here, including actual, implied, possible, potential, and the coordination "bias" or "prejudice," give readers a basis for recovering the intended meaning, and the profile offers lexicography an evidence base for ordering senses by contemporary salience.

4.2 Limitations

The sub-corpora are unequal: the prejudice, statistical, and political-legal datasets support stable statistics, but the bowls, verb, fabric, and preference sets are small, bowls yielding only three lines, so their cues rest on descriptive rather than inferential evidence. The Daily Usage sub-corpus held only 28 instances, 11 resolved. Several contingency-table cells had expected frequencies below five, mitigated but not removed by Monte Carlo simulation, and the bowls row was excluded from Cramér's V. The eight Cambridge senses were collapsed into seven sub-corpora and the specialised electronics sense recorded only in Merriam-Webster was not profiled, the corpora are predominantly American, and the comprehension claims are inferential rather than experimental.

4.3 Future Research Suggestions

The most immediate direction is enlarging the under-represented sub-corpora and the everyday-usage sample, permitting significance testing for the smaller senses. A diachronic study using the Corpus of Historical American English would measure the attenuation and

broadening directly rather than inferring it, embedding-based sense induction could serve as an independent check on the cue-based classification, and the comprehension claims invite psycholinguistic testing of whether lay readers misread the procedural sense.

4.4 Conclusion

In present-day English "bias" is a layered, register-sensitive polysemy network whose senses are lexicographically rooted, distinguishable through their collocates, and unevenly distributed across the registers in which readers encounter the word. The paper thereby contributes to the linguistic understanding of polysemy and to making political and legal communication more comprehensible.

5 Data and Code Availability

The R script reproducing every table and statistical test, the eight sub-corpus profiles, the collocate and cluster tables, the sense-by-cue contingency matrix, and the full cue inventories are openly available at Nguyen and Nguyen (2026), https://doi.org/10.5281/zenodo.21940299. Concordance lines are not redistributed, as the source corpora are licensed, but the query settings needed to regenerate them are deposited with the data.

Acknowledgements

This research was supported by Van Hien University, Vietnam.

References

Anthony, L. (2024). *AntConc* (Version 4.3.1) [Computer software]. Waseda University. https://www.laurenceanthony.net/software/antconc/

Aronson, J. K. (2018, April 10). A word about evidence: 4. Bias: Etymology and usage. *Catalog of Bias*. https://catalogofbias.org/2018/04/10/a-word-about-evidence-4-bias-etymology-and-usag/

Australian Law Reform Commission. (2021). *Cognitive and social biases in judicial decision-making* (Background Paper JI6). https://www.alrc.gov.au/wp-content/uploads/2021/04/JI6-Cognitive-Biases-in-Judicial-Decision-Making.pdf

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

Camacho-Collados, J., & Pilehvar, M. T. (2018). From word to sense embeddings: A survey on vector representations of meaning. *Journal of Artificial Intelligence Research, 63*, 743–788. https://doi.org/10.1613/jair.1.11259

Cambridge University Press. (n.d.). Bias. In *Cambridge Dictionary*. Retrieved May 2026, from https://dictionary.cambridge.org/dictionary/english/bias

Chen, W. F., Al Khatib, K., Wachsmuth, H., & Stein, B. (2020, November). Analyzing political bias and unfairness in news articles at different levels of granularity. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science* (pp. 149–154). https://doi.org/10.18653/v1/2020.nlpcss-1.16

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Curley, L. J., Munro, J., & Dror, I. E. (2022). Cognitive and human factors in legal layperson decision making: Sources of bias in juror decision making. *Medicine, Science and the Law, 62*(3), 206–215. https://doi.org/10.1177/00258024221080655

Davies, M. (2008–). *The corpus of contemporary American English (COCA)*. https://www.english-corpora.org/coca/

Geyh, C. G. (2013). The dimensions of judicial impartiality. *Florida Law Review*, *65*(2), 493–552.

Harper, D. (n.d.). Bias. In *Online Etymology Dictionary*. Retrieved May 2026, from https://www.etymonline.com/word/bias

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524773

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511981395

Merriam-Webster. (n.d.). Bias. In *Merriam-Webster.com dictionary*. Retrieved May 2026, from https://www.merriam-webster.com/dictionary/bias

Mills, L. G. (1999). *Penchant for prejudice: Unraveling bias in judicial decision-making*. University of Michigan Press.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys, 41*(2), 1–69. https://doi.org/10.1145/1459352.1459355

Nguyen, D. H., & Nguyen, N. K. (2026). *Corpus data and R analysis code for a register-based disambiguation study of "bias" in contemporary English* (Version 1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.21940299

Nguyen, D., Doğruöz, A. S., Rosé, C. P., & de Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, *42*(3), 537–593. https://doi.org/10.1162/COLI_a_00258

Nugent, D. C. (1994). Judicial bias. *Cleveland State Law Review*, *42*(1), 1–58.

Oxford University Press. (n.d.). Bias. In *Oxford Learner's Dictionaries*. Retrieved May 2026, from https://www.oxfordlearnersdictionaries.com/definition/english/bias_1

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rädler, P. (1997). Independence and impartiality of judges. In *The right to a fair trial* (pp. 727–746). Springer-Verlag. https://hrlibrary.umn.edu/fairtrial/wrft-rae.htm

Ramsay, A. (2025). *'Partisan bias' is like 'cancer'* (arXiv:2510.05114). arXiv. https://doi.org/10.48550/arXiv.2510.05114

Rodrigo-Ginés, F.-J., Carrillo-de-Albornoz, J., & Plaza, L. (2024). A systematic review on media bias detection: What is media bias, how it is expressed, and how to detect it. *Expert Systems with Applications, 237*, Article 121641. https://doi.org/10.1016/j.eswa.2023.121641

Schane, S. (2006). *Language and the law*. Continuum.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Solan, L. M. (1993). *The language of judges*. University of Chicago Press.

Tiersma, P. M. (1999). *Legal language*. University of Chicago Press.

Weisberg, H. I. (2010). *Bias and causation: Models and judgment for valid comparisons*. Wiley.