UCLA
UCLA Electronic Theses and Dissertations

Title

Text Mining Judicial Dissent: A Computational Analysis of the California Supreme Court

Permalink

https://escholarship.org/uc/item/7hc7s1kz

ISBN

9798265490216

Author

Xu, Runlong

Publication Date

2025-12-11

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

Text Mining Judicial Dissent:

A Computational Analysis of the California Supreme Court

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics and Data Science

by

Runlong Xu

2025

© Copyright by

Runlong Xu

2025

ABSTRACT OF THE THESIS

Text Mining Judicial Dissent:

A Computational Analysis of the California Supreme Court

by

Runlong Xu

Master of Applied Statistics and Data Science

University of California, Los Angeles, 2025

Professor Yingnian Wu, Chair

This thesis investigates how dissenting justices of the California Supreme Court articulate disagreement, and how the nature of that dissent has shifted across two major institutional eras. Through a computational examination of full-text judicial opinions from the Third Series (1969–1991) and Fourth Series (1991–2016), sourced from the Harvard Caselaw Access Project, the study analyzes changes in the rhetorical and emotional architecture of dissent.

Drawing on contemporary Natural Language Processing techniques—including TF-IDF for lexical salience, NRC sentiment scoring for emotional profiling, and Latent Dirichlet Allocation for thematic discovery—the project develops a quantitative account of the minority voice within the court. These tools allow for systematic identification of linguistic trends that traditional

doctrinal analysis often overlooks, revealing how dissenters construct authority, signal disagreement, and frame legal conflict.

The empirical findings suggest a marked rhetorical pivot between the two eras. Dissents in the Cal.3d period are characterized by vocabulary rooted in procedural review and a comparatively optimistic emotional tone. By contrast, dissenting opinions in the Cal.4th period move toward a more statutory and governance-oriented register, accompanied by stronger expressions of negative affect—particularly fear and disgust. These shifts align with broader historical interpretations of the court's transition from an era of expansive judicial reasoning to one more constrained by statutory interpretation and institutional caution.

By offering a reproducible framework for measuring linguistic and emotional changes in judicial writing, this research contributes to the growing literature on computational legal studies. It demonstrates how large-scale text analysis can clarify doctrinal transformations, illuminate judicial identity, and uncover subtle ideological currents within appellate courts. Future work may incorporate transformer-based models for contextual sentiment, explore dissent–majority divergence more explicitly, or extend the methodology to comparative state court systems.

The thesis of Runlong Xu is approved.

Frederic R Paik Schoenberg

Maryam Mahtash Esfandiari

Yingnian Wu, Committee Chair

University of California, Los Angeles

2025

To my parents and friends. . .

who are always by my side to love and support me

My sincere gratitude toward you

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Introduction

The evolution of judicial discourse often mirrors broader societal shifts, with dissenting opinions playing a crucial role in highlighting ideological changes and legal advancements. This study utilizes computational linguistics to analyze dissenting opinions across two eras of the California Supreme Court: the Third Series (Cal.3d, 1969–1991) and Fourth Series (Cal.4th, 1991–2016) of the California Reports. As one of the most influential state courts in the United States, the California Supreme Court has been instrumental in shaping national legal trends, with its dissents serving as key indicators of evolving judicial philosophies. This introduction outlines the significance of the court, establishes the study's research objectives, and emphasizes the contributions of this work to both legal scholarship and computational text analysis.

1.1 Background and Motivation

The California Supreme Court has historically functioned as a primary engine of American legal federalism, particularly during the era of "New Judicial Federalism" that defined the Third Series (*Cal.3d*) of the California Reports [Mil16]. Under the leadership of Chief Justices Donald Wright and Rose Bird, the court aggressively employed the doctrine of "independent state grounds," interpreting the California Constitution to provide broader civil liberties than the federal Constitution, most notably in the volatile area of capital punishment. This judicial philosophy reached its zenith in *People v. Anderson* (1972), where the court declared the death penalty unconstitutional under the state's "cruel or unusual" punishment clause, a decision that stood in

direct defiance of prevailing public sentiment and catalyzed a decades-long conflict between the judiciary and the electorate [GCC20].

The transition to the Fourth Series (*Cal.4th*) was precipitated by the historic 1986 retention election, in which California voters removed Chief Justice Bird and two associate justices, marking a decisive end to the court's liberal hegemony and fundamentally altering its institutional identity [GCC20]. The subsequent era, led by Chief Justices Malcolm Lucas and Ronald George, was characterized by a distinct ideological retrenchment and a shift toward "judicial restraint," particularly regarding the enforcement of the voter-approved "Victims' Bill of Rights" (Proposition 8) [Sch16] and the affirmation of capital sentences. Although the George Court later revived the independent state grounds doctrine in the landmark *In re Marriage Cases* (2008), the *Cal.4th* era largely represents a pragmatic institutional response to the political trauma of 1986, shifting from the "judicial activism" of the Bird era to a more centrist, text-based jurisprudence.

While legal scholars have traditionally analyzed dissents through individual case studies, they have often overlooked broader linguistic trends across hundreds of opinions. Today, computational tools like sentiment analysis and topic modeling provide new opportunities to examine these texts at scale, revealing ideological and rhetorical patterns that manual analysis may miss. Motivated by this potential, this study applies natural language processing (NLP) techniques to explore key features of dissenting opinions: whether dissents became more emotionally charged during the Cal.4th era, how the vocabulary of dissent evolved over time, and the thematic priorities that defined each era's minority opinions. In addition to advancing academic inquiry, this research has practical implications. A deeper understanding of dissenting opinions contributes to ongoing discussions about judicial transparency and persuasive strategies within state courts. By adapting NLP techniques to the formal structure and specialized terminology of legal texts, this study

develops a replicable framework for comparative analysis that can be applied beyond California's borders.

1.2 Research Objectives and Contributions

The primary objectives of this research are as follows: to measure sentiment and emotional tone patterns in dissenting opinions from the Cal.3d and Cal.4th eras using NRC lexicon analysis [MT13], to identify distinctive vocabulary and thematic patterns through TF-IDF keyword extraction [LWZ18] and LDA topic modeling [JWF18], and to analyze the relationship between justices' ideological positions and their linguistic choices in dissenting opinions.

The contributions of this work are both theoretical and practical. For California's legal community, this study provides empirical evidence of the rhetorical evolution in dissenting opinions during the court's ideological shift, offering quantitative insights to complement traditional doctrinal analysis. For legal scholars, this research establishes measurable patterns of judicial polarization that are expressed through language, enriching the understanding of dissent as both a legal and rhetorical tool.

From a methodological standpoint, this study makes two key advances. First, it adapts NLP techniques to the unique challenges posed by legal texts, including the embedded citations, formal structure, and specialized terminology inherent in judicial opinions. Second, it creates a analytical framework for comparing judicial discourse across state supreme courts, which can be applied to future studies comparing dissenting opinions from various jurisdictions. These innovations are particularly timely given the growing interest in computational legal studies and the increasing importance of data-driven legal analysis.

CHAPTER 2

Legal Context

This chapter provides the legal-historical framework necessary for analyzing dissenting opinions within the California Supreme Court's Cal.3d (1969–1991) and Cal.4th (1991–2016) eras. It outlines the court's national influence, the ideological shift between the two periods, and the role of dissenting opinions as both instruments of legal development and as rhetorical contestations of majority decisions. This contextual foundation will ground the computational study, highlighting how dissenting language reflects institutional identity and doctrinal conflict within the California Supreme Court.

2.1 The California Supreme Court's Judicial Role

The California Supreme Court serves as the state's court of last resort, exercising discretionary review over the Courts of Appeal and issuing binding precedent on all state courts. Its docket—especially in areas such as constitutional interpretation, tort innovation, and regulatory law—positions it as a major architect of California's legal order. Scheiber's institutional history emphasizes that the Court has long shaped the balance between legislative authority, direct democracy, and individual rights, reflecting California's unusually dynamic constitutional environment [Sch16].

Nationally, the Court is regarded as one of the most influential state supreme courts. Empirical citation studies show that California decisions, particularly in products liability, consumer protection, and market-share liability, have been among the most frequently followed by other

jurisdictions, giving the Court an outsized role in the development of American common law. This "leader court" status means that doctrinal shifts in California often diffuse beyond state borders.

The Court's work is also embedded in a political context shaped by direct democracy and partisan appointment patterns. Miller argues that the Court engages in an ongoing "dialogue" with the electorate through its interpretation of initiatives and tax-limitation provisions, making it a key institution mediating popular will [Mil16]. More recent quantitative analysis by Gergen, Carrillo, Chen, and Quinn documents systematic ideological alignments in closely divided cases, indicating that partisan appointment groups shape a meaningful share of modern outcomes [GCC20].

For this thesis, these institutional features matter because dissenting opinions are a primary venue where justices articulate competing visions of constitutional meaning, statutory interpretation, and the Court's broader role. Across both the Cal.3d (1969–1991) and Cal.4th (1991–2016) eras, dissents reveal internal doctrinal conflict and help trace how the Court's identity evolves in response to political, institutional, and jurisprudential pressures.

2.2 Significance of the Cal.3d and Cal.4th Eras

The Cal.3d period, dominated by the tenure of Chief Justices Donald Wright and Rose Bird, was defined by the aggressive application of "independent state grounds." Under this doctrine, the court interpreted the California Constitution to provide broader civil liberties than the federal Constitution, particularly in the wake of the U.S. Supreme Court's conservative turn. The era reached its ideological zenith—and its political breaking point—with *People v. Anderson* (1972), where the court declared the death penalty unconstitutional under the state's "cruel or unusual" punishment clause [And72]. This "judicial activism," while academically influential, placed the

court in direct conflict with the electorate, leading to the unprecedented removal of Chief Justice Bird and two associate justices in the 1986 retention election.

The subsequent era, led by Chief Justices Malcolm Lucas and Ronald George, was characterized by institutional stabilization and judicial restraint. The electorate's passage of Proposition 8 (The Victims' Bill of Rights) constitutionally abrogated the court's ability to use independent state grounds to exclude criminal evidence, forcing California courts to follow federal interpretations of the Fourth Amendment [Lan85]. Consequently, the Cal.4th era saw a sharp decline in the reversal of criminal convictions and a general retreat from the creation of new tort liabilities. However, the court did not abandon its independence entirely; in the landmark *In re Marriage Cases* (2008), the George Court utilized independent state grounds to extend strict scrutiny equal protection analysis to sexual orientation, striking down the state's ban on same-sex marriage [Mar08].

2.3 Dissenting Opinions in State Jurisprudence

Dissenting opinions in state supreme courts serve as vital instruments of doctrinal development and institutional expression. Though lacking precedential force, they preserve competing legal interpretations that may influence future rulings or legislative reform. Dissents also function rhetorically, offering justices a venue to articulate principled disagreement and frame alternative constitutional or statutory visions.

Empirical research shows that dissenting behavior often reflects ideological polarization and institutional context. Courts with discretionary dockets, like the California Supreme Court, tend to produce more dissent in high-stakes or politically salient cases, where internal divisions are sharper

and doctrinal stakes are higher [ELP11]. In such settings, dissent becomes a mechanism for contesting majority reasoning and registering jurisprudential dissent.

Dissents also occupy a distinct place in legal discourse. They are situated in what scholars call a "hybrid" position—neither law nor mere commentary—aimed at shaping future understanding rather than immediate outcome [Fro24]. This status enables dissenting justices to write with greater rhetorical freedom, targeting both internal court audiences and external legal and political actors.

The long-term significance of dissent is particularly evident in the California Supreme Court's history. Justice Jesse W. Carter's prolific dissents, often ahead of their time, later formed the foundation for doctrinal changes in areas like search and seizure, due process, and evidentiary standards [Opp10]. His legacy illustrates how dissenting opinions can crystallize alternative legal frameworks that gradually gain institutional traction.

CHAPTER 3

Data Construction

3.1 Data Collection and Source Overview

The dataset used in this study was sourced from the Harvard Caselaw Access Project (CAP) [Cap], a large-scale open-access archive of U.S. court decisions. Each case is provided in structured JSON format, including full-text judicial opinions and rich metadata. To explore linguistic and rhetorical shifts in dissenting opinions of the California Supreme Court, this project focuses on two major publication eras: Cal.3d (1969–1991) and Cal.4th (1991–2016).

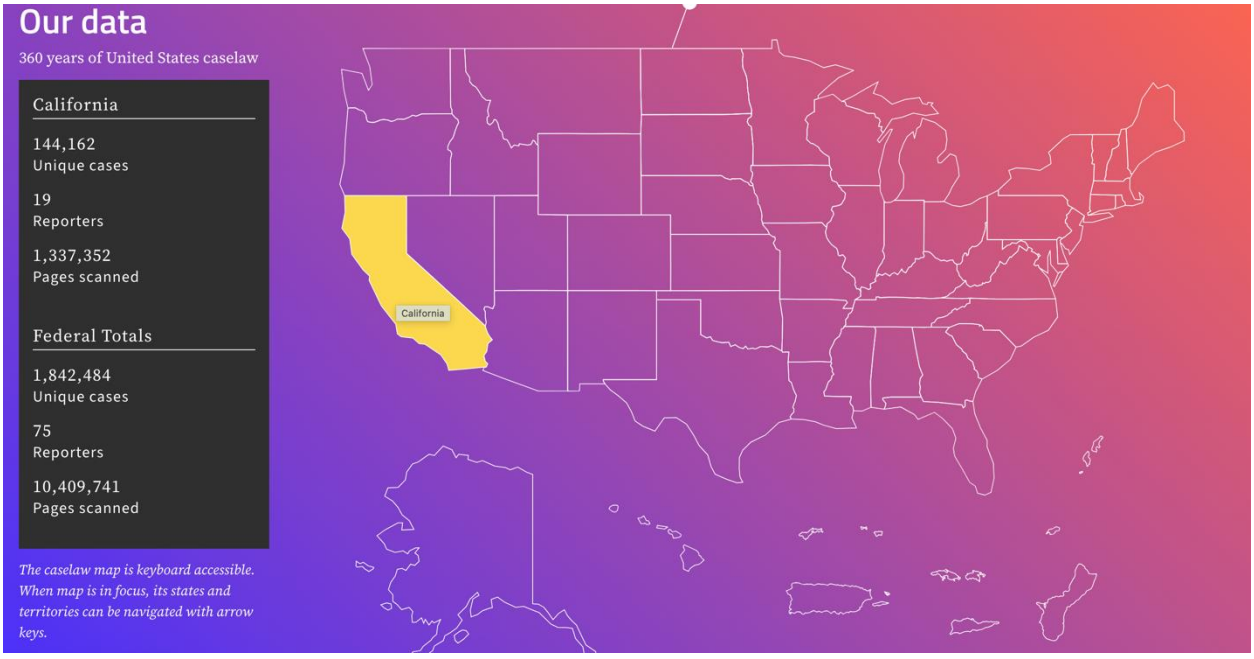

Figure 3.1: California Law Cases Overview

From the CAP repository, a stratified sample of case files was selected from across both eras. Volumes were chosen at approximately five-year intervals to ensure temporal balance and representation of different judicial administrations. This sampling strategy allows for meaningful comparison of dissenting rhetoric over nearly five decades without overrepresenting any single period.

Each JSON file represents an individual case and contains key elements such as the case title, decision date, court metadata, and full-text opinions. Of particular importance is the casebody.opinions array, which contains all written opinions, including dissents, majority opinions, and their respective authors.

All data were downloaded and processed locally. The JSON structure allows for efficient extraction of dissenting opinions for natural language processing (NLP) tasks such as sentiment analysis, topic modeling, and lexical comparison [KK24]. While the case files contain both majority and dissenting texts, the primary analytical focus of this thesis remains on dissenting opinions. However, majority texts may be referenced for contextual interpretation or comparative baselining when needed.

3.2 Opinion Extraction & Cleaning Pipeline

The Opinion Extraction & Cleaning Pipeline forms the cornerstone of this study's text mining process. Using data sourced from the Harvard Caselaw Access Project (CAP), this pipeline was designed to systematically extract, preprocess, and clean dissenting opinions from the California Supreme Court's decisions. The extracted data underwent a multi-step process, ensuring that the final dataset was optimized for in-depth linguistic analysis, such as sentiment analysis, topic modeling, and lexical comparisons.

The first step in the pipeline is the extraction of relevant data from the structured JSON format used by CAP. Each case in the CAP repository is represented by a JSON file containing rich metadata about the case, including the case name, docket number, decision date, and full-text opinions. A key component of each file is the *casebody.opinions* array, which contains the full-text opinions for each case, categorized by type (e.g., majority, dissent, concurring).

For the purposes of this study, the focus was primarily on dissenting opinions, which offer unique insights into judicial disagreement and the rhetorical strategies employed by justices. In addition to the opinion text, important metadata such as the decision date and case name were also extracted for contextual analysis. This metadata is essential as it enables the tracking of shifts in dissenting rhetoric over time, across different judicial administrations.

Once the dissenting opinions were extracted, they underwent a tokenization process, where the text was split into individual tokens, typically words. Tokenization is a fundamental step in natural language processing (NLP), as it allows the text to be reduced to its core components for subsequent analysis [Cha23]. The aim of tokenization was to break down the unstructured opinion text into smaller, manageable units of analysis which are words or phrases.

Following tokenization, the text cleaning process began. The objective here was to remove any noise or irrelevant elements from the text that would otherwise skew the analysis [Cha23]. Several steps were involved in this process:

1. Removal of Non-Alphabetic Characters: The text was filtered to remove tokens that consisted solely of numeric characters or punctuation marks. These elements do not provide useful information for linguistic analysis and could disrupt the analysis of the language used in dissenting opinions.

2. Filtering of Short Words: Words with fewer than three characters were excluded, as these typically represent stopwords or irrelevant tokens (e.g., "the", "and", "of"). This step was essential in ensuring that only meaningful words contributed to the analysis.

3. Stopword Removal: A stopword removal procedure was applied to filter out high-frequency words that do not carry significant meaning in the context of the analysis. This includes common words such as "court", "judge", and "law". Additionally, a set of custom stopwords was created specifically for this legal corpus, addressing terms that frequently appear in judicial opinions but do not contribute substantially to the analysis of dissenting language.

4. Conversion to Lowercase: All tokens were converted to lowercase to standardize the data. This step ensures consistency by eliminating case-based duplications of words (e.g., "Case" vs. "case").

5. Retention of Alphabetic Tokens: Finally, tokens that contained non-alphabetic characters (e.g., numbers or symbols) were discarded. This step was particularly important for maintaining the quality of the dataset, focusing solely on meaningful words for analysis.

After these steps, the dataset was transformed from a large body of raw text into a cleaned, structured dataset consisting of individual tokens. Each dissenting opinion was now represented by a list of tokens, with additional metadata such as doc_id, volume, file, opinion_type, and decision_date retained for analysis. The final dataset consisted of the following columns:

- **doc_id**: A unique identifier for each case, combining the volume and file name.
- **volume**: The volume or publication era (e.g., Cal.3d or Cal.4th).

- **file**: The specific collection in which a legal report can be found.
- **opinion_type**: The type of opinion (e.g., majority, dissent).
- **tokens**: A list of tokenized words that form the cleaned and processed text of the dissenting opinion.
- **decision_date**: The date the decision was made.
- **name_abbreviation**: The abbreviated name of the case.

This cleaned and tokenized dataset served as the basis for the subsequent natural language processing tasks. The tokens provided the input for advanced analyses, such as sentiment analysis, where the emotional tone of dissenting opinions could be assessed, and topic modeling, where common themes and subjects across various dissenting opinions could be identified.

CHAPTER 4

Methodology

This section details the core analytical methods employed in this thesis to examine the linguistic, emotional, and thematic patterns in dissenting opinions from the California Supreme Court. The primary methods utilized are Sentiment Analysis (using the NRC Emotion Lexicon), Term Frequency-Inverse Document Frequency (TF-IDF) for keyword extraction, and Latent Dirichlet Allocation (LDA) topic modeling. These techniques were selected specifically for their capacity to reveal subtle yet meaningful shifts in judicial rhetoric, tone, and thematic emphasis across the Cal.3d and Cal.4th eras [BDI18].

4.1 Sentiment Analysis

Sentiment analysis is an established method for evaluating the emotional content embedded within textual data. By categorizing words according to emotional attributes, this technique enables researchers to systematically capture and quantify the emotional undercurrents of written opinions. In this thesis, sentiment analysis was implemented using the NRC Emotion Lexicon—a comprehensive, widely used dictionary that classifies terms into distinct emotional categories, including anger, fear, sadness, joy, surprise, trust, as well as broader positive and negative sentiment orientations [MT13].

Each dissenting opinion was tokenized into words and matched against the NRC lexicon, generating scores for various emotional categories. By aggregating these scores for each opinion

and normalizing them relative to opinion length, it becomes possible to detect overarching emotional trends, such as increases in expressions of anger or trust across different judicial periods. This approach highlights not only how dissenting justices communicate emotional stances when disagreeing with the majority but also provides valuable insights into whether and how emotional rhetoric has evolved between the Cal.3d and Cal.4th eras.

Beyond merely identifying the emotional tone of dissents, this analysis can uncover potential divergences in emotional emphasis relative to majority opinions. As such, sentiment analysis provides a nuanced, quantitative lens through which to explore the judiciary's rhetorical dynamics and emotional framing of critical legal debates.

4.2 TF-IDF for Keyword Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) is a powerful statistical measure commonly employed to determine how distinctive or characteristic a word is within a document and is a core part of statistical NLP frameworks applied to legal texts [Nay24]. Essentially, TF-IDF highlights words that appear frequently in specific documents (high term frequency) yet are relatively uncommon across the corpus as a whole (high inverse document frequency), indicating their importance or uniqueness to those documents [LWZ18].

Formally, for a given word *t* in a dissenting opinion *d*, the TF-IDF score is calculated as follows:

$$\text{TF-IDF}(t, d) = \text{tf}(t, d)\ \log\left(\frac{N}{n_t}\right) \tag{4.1}$$

Where $\text{tf}(t, d)$ is the frequency of the term t in the document d, N is the total number of dissenting opinions, and n_t is the number of dissenting opinions containing the term t. Two light, content-neutral gates keep the rankings stable and readable: a token must appear in at least three different

dissents and at least five times overall. After scoring, duplicate *(doc_id, word)* rows are collapsed by taking the mean TF–IDF of each word across all dissents where it appears; that mean becomes the word's corpus-level salience. Computing IDF within dissents only keeps the focus squarely on dissenting rhetoric rather than blending it with majority style. The outcome is a refined set of keywords that foregrounds persistent doctrinal and procedural terms (rather than one-off names), making it easier to read shifts in legal emphasis over time and to link vocabulary choices to broader rhetorical strategies.

4.3 LDA Topic Modeling

Latent Dirichlet Allocation (LDA) is a topic modeling technique that utilizes machine learning that discovers latent thematic structures in a corpus. LDA assumes that each document (or in this case, each dissenting opinion) is a mixture of several topics, and each topic is characterized by a distribution of words. This method enables the identification of themes or topics that recur across dissenting opinions, so that we can discover the evolution of legal thought and ideological shifts over time [DTA24].

In this thesis, LDA was employed to analyze the dominant themes within the dissenting opinions of the California Supreme Court across two major publication eras (Cal.3d and Cal.4th). This approach allows for the detection of shifting ideological patterns, such as the rise of certain legal arguments or the shift in focus from procedural issues to more rights-based or societal issues. For instance, LDA may uncover topics like due process, equality, or freedom of speech that evolve over time, providing insights into how legal discourse and social issues influence judicial rhetoric.

Moreover, LDA facilitates a deeper understanding of how individual justices align with or diverge from dominant judicial trends. The topics identified through LDA not only reflect the legal issues

that dominate dissenting opinions but also reveal the underlying ideological divides within the court across different time periods.

4.4 Era Comparison Framework

The analysis in this thesis is structured around a comparative framework that contrasts dissenting opinions from two key periods of the California Supreme Court: The Cal.3d (1969–1991) era and the Cal.4th (1991–2016) era. This comparison allows for the examination of shifts in legal rhetoric, emotional tone, and thematic focus across time, revealing how changes in societal and political contexts may have influenced the Court's dissenting opinions.

The focus of the era comparison is on the dissenting opinions, as these often reflect shifts in ideological perspectives and legal challenges. By examining the language, emotional tone, and topics of dissenting opinions in both periods, this framework aims to uncover whether the California Supreme Court's dissenting voices have become more or less assertive, emotional, or ideologically distinct over time.

To facilitate comparison, the core techniques—sentiment analysis, TF-IDF, and LDA topic modeling—were applied to both sets of dissenting opinions. Sentiment analysis is used to gauge shifts in emotional expression, while TF-IDF highlights changes in legal terminology. LDA topic modeling provides insight into the evolution of themes and issues addressed by dissenting justices. The results of these analyses are then compared across the two eras, allowing for a nuanced understanding of how the Court's dissenting opinions have evolved in response to changing legal and social contexts

CHAPTER 5

Lexical Shift

This chapter traces how the vocabulary of California Supreme Court dissents changed from the Cal.3d (1969-1991) to the Cal.4th (1991-2016) series. It moves in three steps. First, opinion-level TF-IDF surfaces the hundred tokens that give each era its distinctive lexical "accent." Second, bigram network graphs reveal which of those tokens habitually travel together, exposing the short phrases that carry doctrinal meaning. Third, a weighted log-odds comparison identifies the individual words that lean most strongly toward one era or the other after controlling for overall frequency. Recent research has demonstrated that topic-sensitive lexical mining techniques, such as TF–IDF and bigram clustering, can effectively extract legal narratives from police and court documents, confirming their relevance for judicial analysis [BMB24].

5.1 Distinctive Dissent Vocabulary by Era

This section reports the vocabulary that most distinctively characterizes dissenting opinions in each era, using TF–IDF computed at the level of a single dissent and then averaged across dissents. Figure 5.1 (Cal.3d, 1969–1991) and Figure 5.2 (Cal.4th, 1991–2016) plot the top 100 tokens by mean TF–IDF. Line length shows salience; the small number to the right of each token is how many different dissents use it. Each "lollipop" ranks a token by its mean TF–IDF within that era. Averaging across documents makes the scores robust to opinion length—long dissents don't win by sheer word count—and the right-hand counts help tell broad, era-defining vocabulary from

more specialized but signature terms. Proper nouns that function as doctrinal shorthands may still appear; here they serve as compact labels for recurring standards rather than noise.

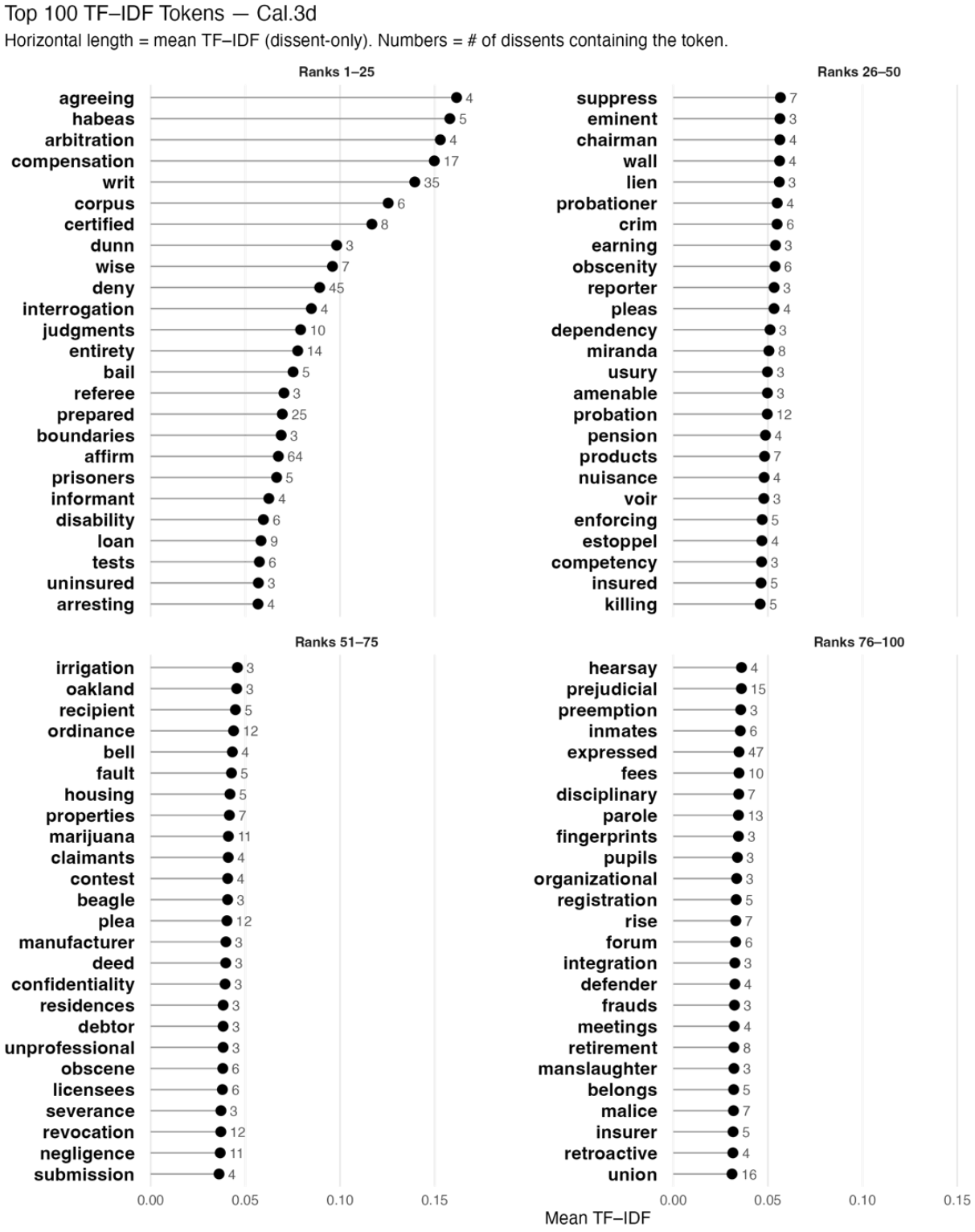

Figure 5.1: Top 100 TF-IDF Tokens (Cal.3d)

As Figure 5.1 shows, Cal.3d dissents lean hard into criminal procedure and post-conviction review—*habeas, corpus, writ, interrogation, miranda, suppress, informant, bail, probationer, voir*. These aren't just generic court words; they're the vocabulary of how cases were decided, from exclusionary rules and voluntariness to the scope of collateral review. A second band of terms—*compensation, judgments, estoppel, nuisance, manslaughter, negligence*—plus institutional markers like *ordinance, registration, fees, housing, inmates,* and *disciplinary* points to steady skirmishes over remedies, public-employment benefits, and regulatory context. The breadth counts next to each token matter: items like *writ, corpus,* and *habeas* show up across many dissents, so this isn't one or two oddball cases driving the picture. And because the ranking uses mean TF–IDF, long opinions don't float to the top by word count alone—distinctive, argument-shaping terms (e.g., *suppress, interrogation, informant*) outrank blander nouns like *probation* or *plea*.)

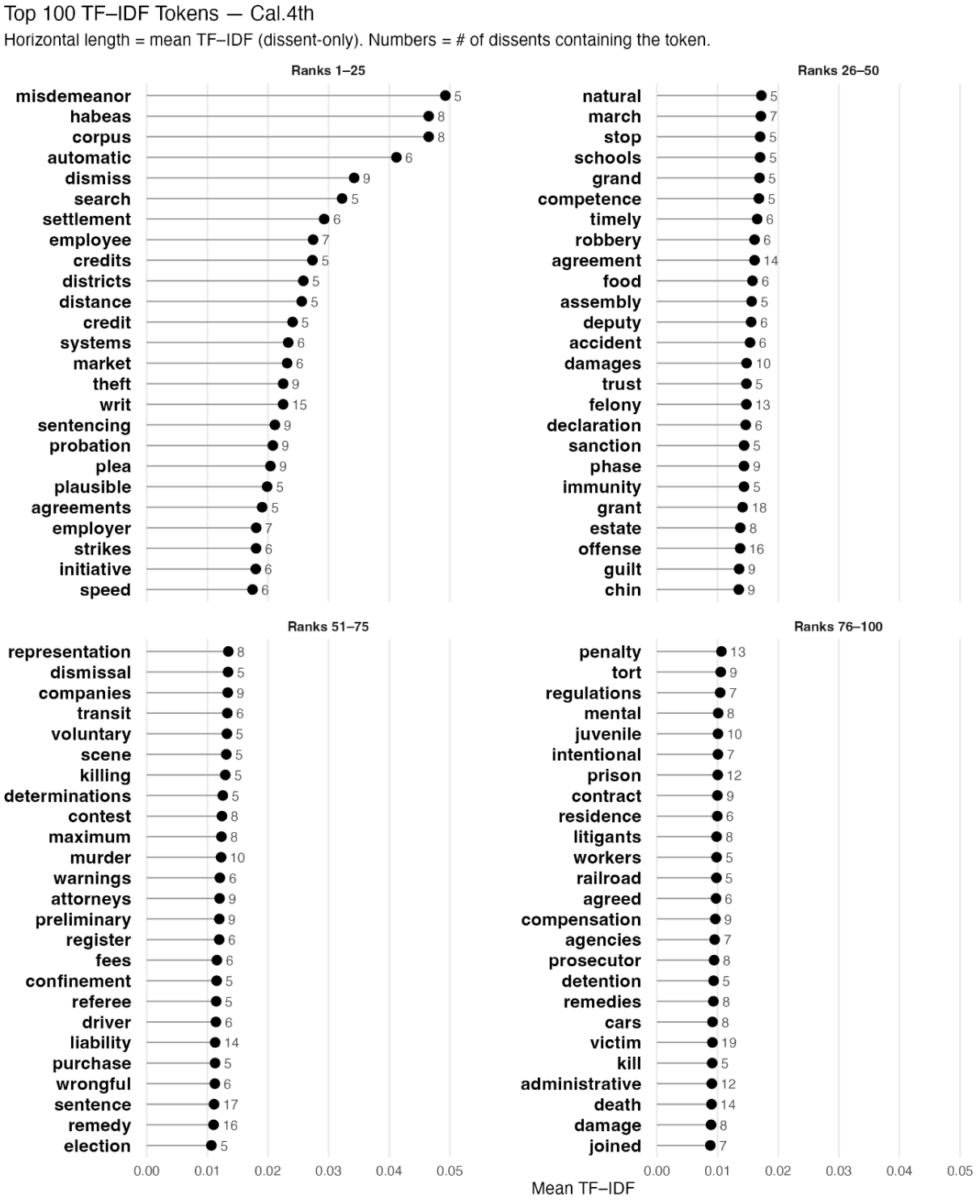

Figure 5.2: Top 100 TF-IDF Tokens (Cal.4th)

Figure 5.2 tilts in a different direction. Cal.4th dissent vocabulary moves toward charging and sentencing frameworks and the statutory/initiative architecture around them: *misdemeanor, plea, sentencing, strikes, wobbler, indictment, vacate, probation* sit at the core. Around that core is a thicker statutory/organizational layer—*registration, initiative, districts, assembly, warn, timely*—and more economic/workplace language—*arbitration, settlement, corporate, employer, employee, consumer, market, systems*.

Read side by side, the two plots suggest a drift from Cal.3d's emphasis on procedural safeguards and remedies toward Cal.4th's emphasis on standardized charging/sentencing regimes and statutory governance. The "procedural-review cluster" in Cal.3d—*habeas/corpus/writ, interrogation, suppress, informant, bail*—remains present in Cal.4th but is joined or overtaken by *misdemeanor, plea, sentencing, strikes, wobbler,* and *vacate*, the vocabulary of a system concerned with structured penalties and post-trial management. On the civil side, *compensation* and related remedial terms give way to *arbitration, settlement, corporate, employee/employer,* and *consumer,* reflecting the period's case mix and the Court's docket composition. Meanwhile, tokens like *initiative, registration, districts,* and *assembly* underline how frequently dissent in the later era is framed as a critique of statutory text, voter-enacted rules, or institutional competence.

5.2 Bigram Networks

While single-word TF–IDF scores tell us which terms dissenting justices leaned on, bigram analysis shows how they linked those terms into recurring phrases. A bigram [TWL02] is simply two consecutive words that appear more often together than chance would suggest—mini-idioms that carry doctrinal shortcuts or stock argumentative moves. After removing boiler-plate and low-value function words, the forty most frequent pairs in each era reveal the connective tissue of

dissent. The resulting networks plot words as nodes and bigrams as arrows; thicker, darker edges mark phrases that recur in dozens of separate opinions.

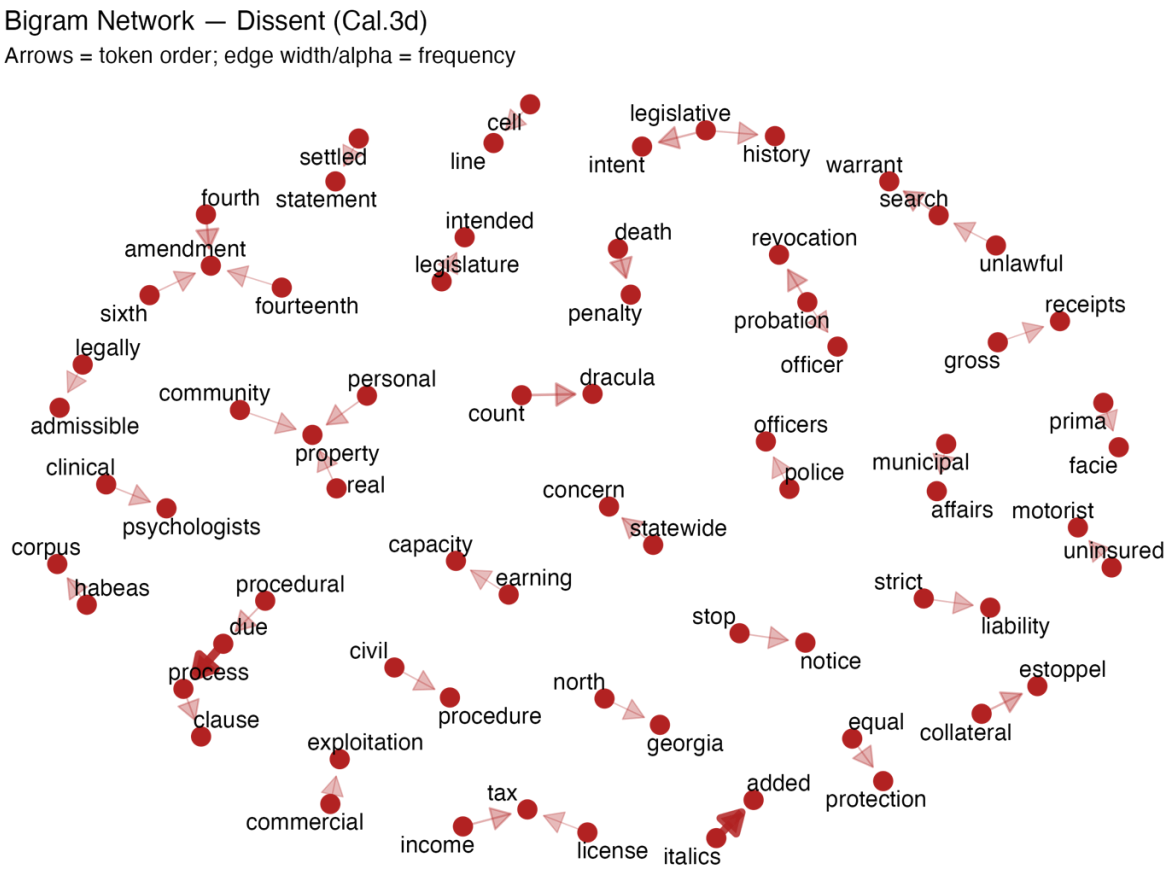

Figure 5.3: Bigram Network – Dissent (Cal.3d)

During the Cal.3d period the network is anchored by a constitutional spine. Pairs such as *fourth amendment*, *sixth amendment*, *fourteenth amendment*, *search warrant*, *habeas corpus*, and *probation revocation* dominate the criminal-procedure half of the graph. Their prominence mirrors the Court's late-1960s and 1970s battles over exclusionary rules, confrontation rights, and proportional review in capital cases. On the civil side the dissenting vocabulary couples remedial doctrines with liability theories: *collateral estoppel*, *strict liability*, *community property*, *inverse condemnation*, *uninsured motorist*. These combinations point to a docket that simultaneously

expanded tort remedies and refined family-property rules. Phrases such as *legislative intent* and *legislative history* show dissenters pushing the majority on purposive statutory reading, while *statewide concern* and *municipal affairs* signal recurring disputes over the reach of home-rule powers. Taken together, the Cal.3d graph portrays a minority voice that drew equal energy from rights-protective criminal procedure and remedy-oriented civil doctrine, making frequent appeals to constitutional text as well as equitable outcomes.

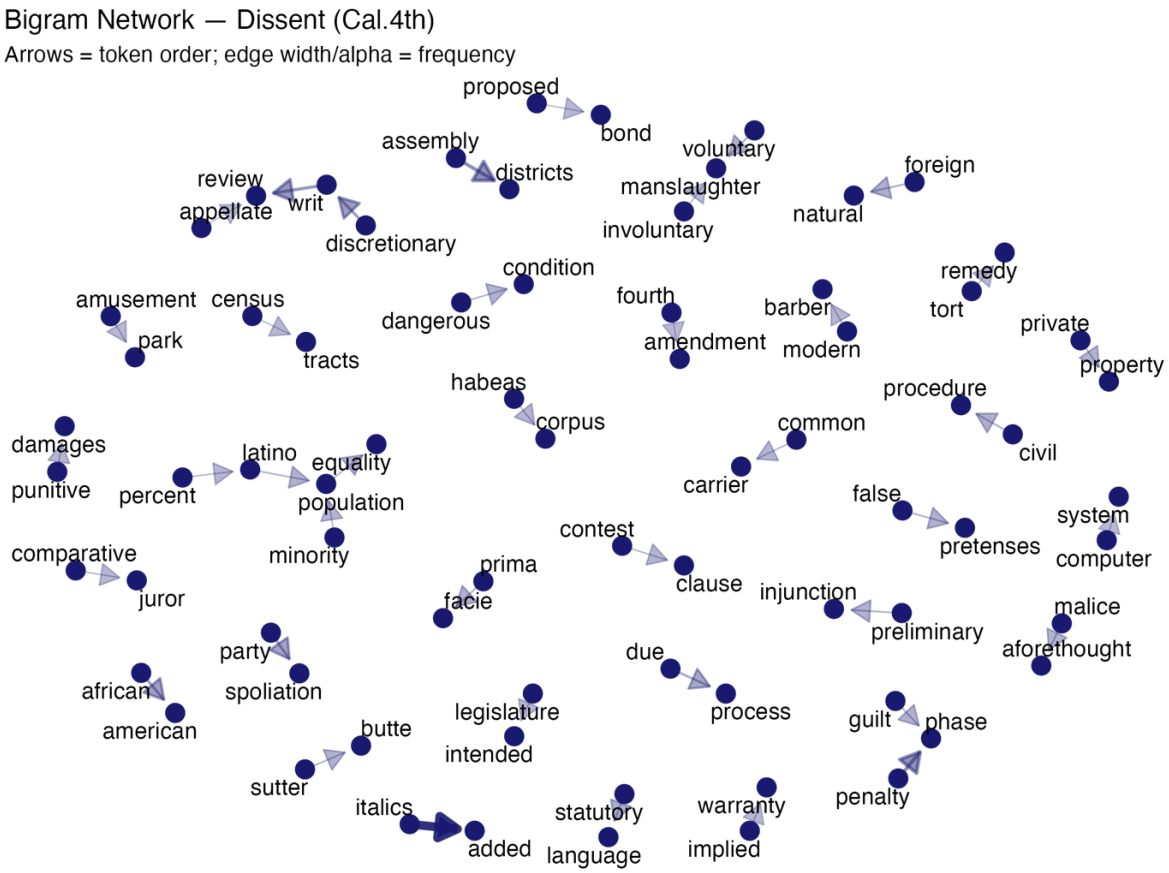

Figure 5.4: Bigram Network – Dissent (Cal.4th)

By contrast, the Cal.4th network pivots toward statutory precision and demographic equality. One prominent cluster joins pair like *assembly districts*, *population equality*, *census tracts*, *african american*, *latino population*, and *minority population*, capturing the Court's intensified

engagement with Voting Rights Act disputes and post-1990 redistricting challenges. A second spine centres on sentencing vocabulary: *penalty phase*, *guilt phase*, *voluntary manslaughter*, *discretionary writ*, and *writ review* reflect the era's California's Three Strikes law enacted in 1994 and capital-punishment jurisprudence. Surrounding these hubs is a layer of modern liability and contract language—*punitive damages*, *dangerous condition*, *implied warranty*, *computer system*—underscoring the Court's heavier consumer-protection and technology-related docket. Bigram pairs such as *statutory language* and *legislature intended* appear far more often than in Cal.3d, confirming a shift toward text-centred critique and initiative-driven governance.

Overall, the Cal.3d network is steeped in constitutional clauses paired with procedural safeguards, reflecting a court still refining post-Warren-era protections. Cal.4th dissenters, operating in a statutory and initiative-heavy landscape, couple demographic terms with sentencing jargon and finely grained statutory references.

5.3 Keyword Contrast Through Log-Odds Lens

The single-token ranks in section 5.1 tell us what dissenters talked about; the bigram maps in section 5.2 show how they stitched concepts together. A third question remains: which vocabulary is truly characteristic of one era rather than the other? To answer it we move from simple frequency to a weighted log-odds comparison that pits the Cal.3d corpus against the Cal.4th corpus as rival "classes."

For every lower-cased token that appears at least 30 times across all dissents, we compute

$$\log\left(\frac{(n_{w,\text{cal3d}}+\alpha)/(N_{\text{cal3d}}+K\alpha)}{(n_{w,\text{cal4th}}+\alpha)/(N_{\text{cal4th}}+K\alpha)}\right)-\log\left(\frac{(n_{-w,\text{cal3d}}+\alpha)/(N_{\text{cal3d}}+K\alpha)}{(n_{-w,\text{cal4th}}+\alpha)/(N_{\text{cal4th}}+K\alpha)}\right) \qquad (5.1)$$

where $n_{w,\text{era}}$ is the token count in that era, N_{era} is the total token count, K is the vocabulary size, and $\alpha = 0.5$ is an uninformative Dirichlet prior [MCQ08]. The prior stabilizes estimates for low-frequency words; larger absolute values signal a stronger association with one era. In the mirrored bar-plot below, positive scores extend right for Cal.3d, and negative scores extend left for Cal.4th.

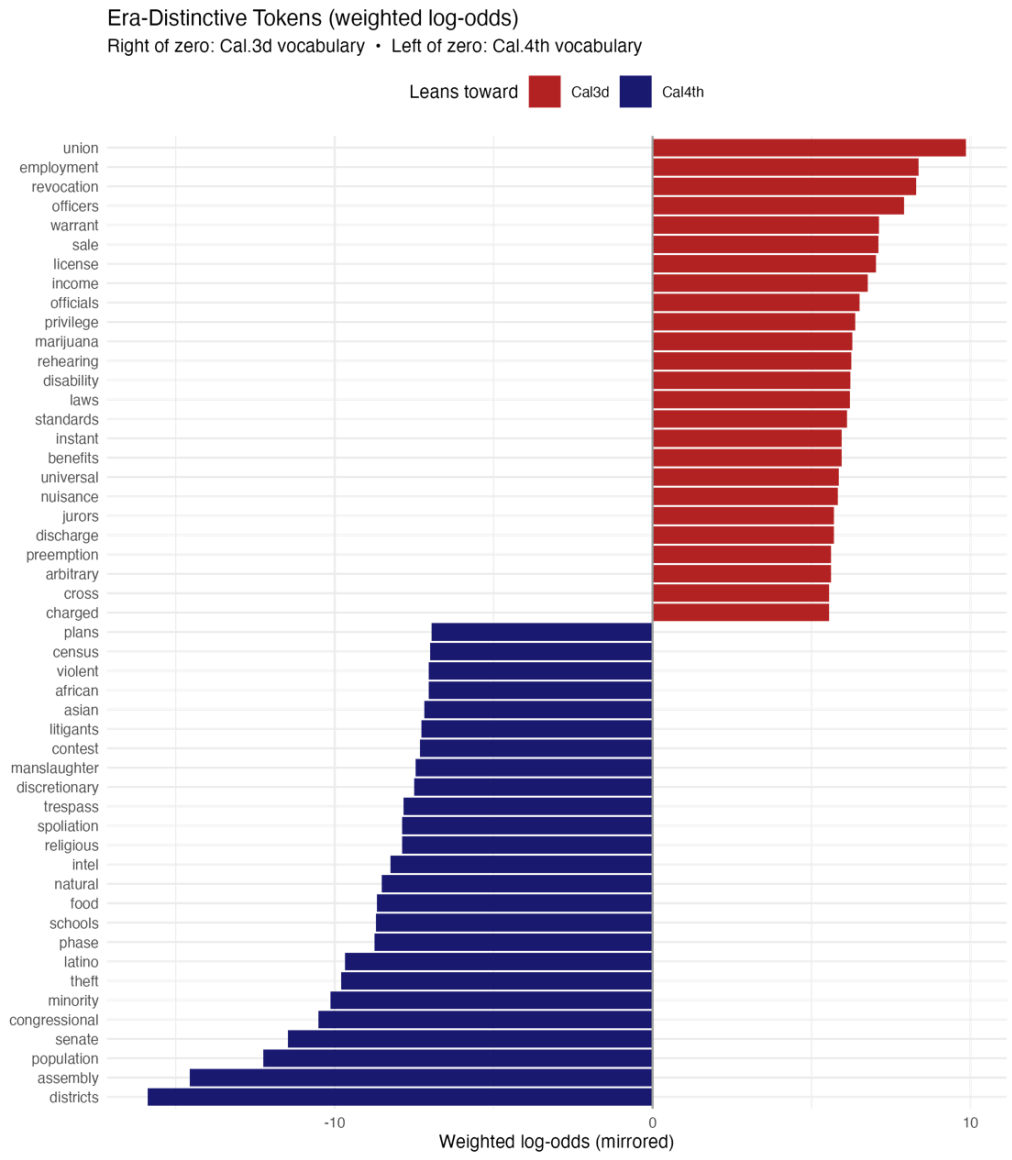

Figure 5.5: Era-Distinctive Tokens (weighted log-odds)

The red half of Figure 5.5 indicates Cal.3d is led by *union, employment, revocation, officers, warrant, license*—a cluster that reads like the mid-century docket: public-employee discipline, labour-management disputes, and police-search challenges. *Income, privilege, marijuana, nuisance* track the era's willingness to test the limits of taxation, evidentiary privilege, and contraband regulation, while *rehearing, standards, preemption* mark procedural skirmishes in which dissenters accused the majority of short-circuiting review or ignoring federal overlaps. Notably, tokens such as *marijuana* and *warrant* echo the bigram spine of *search warrant / unlawful search* that we saw in section 5.2, tightening the picture of a court wrestling with Fourth-Amendment style protections under state law.

The blue bars show that Cal.4th swing farthest left for *districts, assembly, population, senate, congressional, minority, latino, asian, african*—the lexicon of post-1990 redistricting fights and Voting-Rights-Act challenges. They sit beside *theft, phase, manslaughter, discretionary, violent,* tokens that track a sentencing-heavy criminal docket arcing from Three-Strikes to penalty-phase jurisprudence. *Intel, spoliation, trespass, food, natural* signal intellectual-property and product-liability questions related to the technology and consumer-protection lawsuits of the 1990s and 2000s.

The result is sharper: some high-TF-IDF tokens drop out because they are common to both periods, while lower-frequency but era-telling words shoot to the top. Together, the three lexical views reveal a consistent story: Cal.3d dissenters still fought over criminal-procedure safeguards, but their signature vocabulary tilts toward *labour, licensing,* and the procedural mechanics of review—mirroring a court inclined to expand remedies and public-employee rights. Cal.4th dissenters speak the language of *redistricting, sentencing,* and *statutory technicalities,* reflecting a bench preoccupied with initiative-driven statutes, demographic equality, and the collision of old doctrine

with a modern economy. Thus, weighted log-odds supplies the final layer: it shows what dissenters emphasised as well as which words became era markers [MCQ08].

CHAPTER 6

Sentiment & Emotion Trends

Chapter 6 maps how the tone of California Supreme Court dissents evolved from 1969–2016. Using the NRC Emotion Lexicon [MT13], each opinion is converted to per 1000 token rates of positive/negative words and the eight basic emotions, computed at the document level to avoid length bias. We first contrast overall valence across eras (Cal.3d vs Cal.4th), then place the polarity measure on a yearly timeline with uncertainty and the 1991 boundary, and finally decompose that signal into emotion-specific trajectories to see which registers—trust versus fear, sadness, disgust, and others—drive the shifts. The aim is to pinpoint not only whether tone changed, but when it changed and which emotions underwrote those moves.

6.1 Dissent Polarity Shifts

To track how the tone of California Supreme Court dissents changed across eras, Figure 6.1 reports the density of words tagged positive or negative by the NRC Emotion Lexicon—a curated dictionary widely used in text analysis that labels common English words by valence (positive = approval/benefit; negative = criticism/harm/cost). After standard preprocessing, tokens from each dissent were matched to those lists, converted to rates per 1,000 tokens at the opinion level, and then averaged by era. Very short dissents were excluded to stabilize rates, so each bar reflects the mean opinion-level rate, not raw totals.

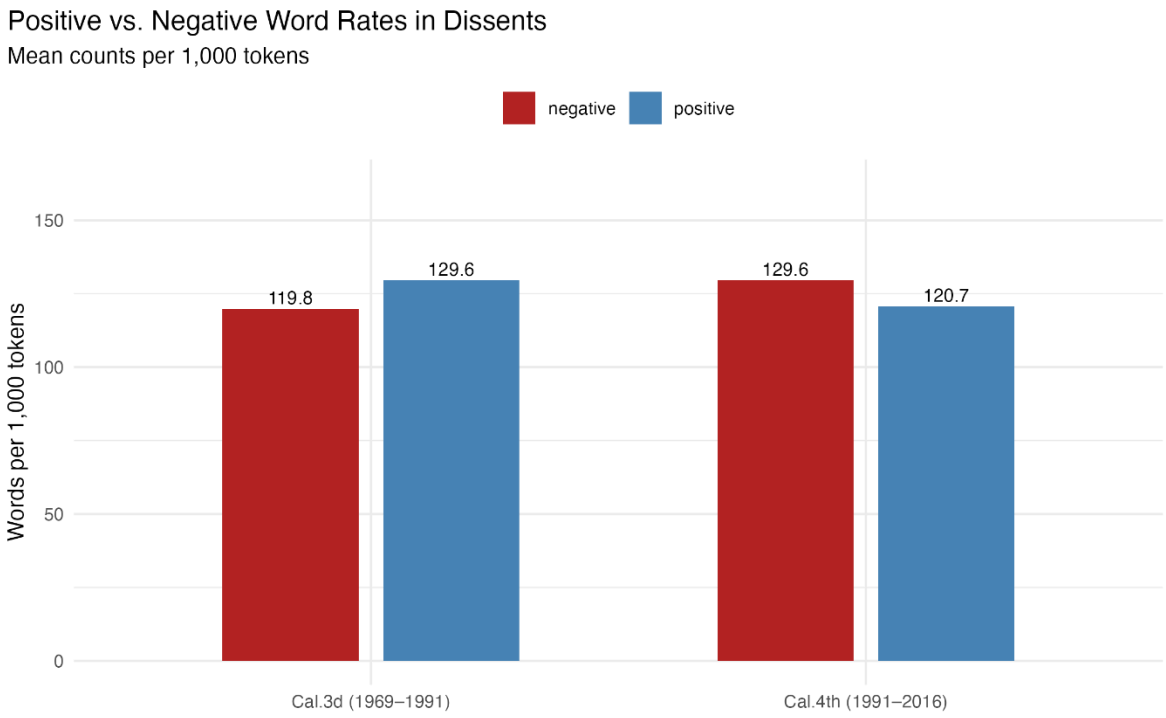

Figure 6.1: Positive vs. Negative Word Rates in Dissents

Read the plot as a crossover. In Cal.3d, positive terms edge out negative ones; in Cal.4th, the ordering reverses, with negative terms slightly more common than positive. The typical Cal.3d dissent contains a little more affirming language than problem-finding language, while the typical Cal.4th dissent contains a little more problem-finding than affirming language. Because each bar reflects an average of opinion-level rates (not raw totals), the pattern should be read as characteristic of the median dissent rather than driven by a few unusually long or heated opinions.

Several features of the plot are worth reading closely. First, the difference between the positive and negative bars within each era is smaller than the difference between eras—i.e., both eras use *both* valences heavily, but the balance flips. Second, the absolute heights show that either valence appears roughly 120–130 words per 1,000—about 12–13% of the vocabulary after cleaning—which underscores how pervasive evaluative language is in dissents. Finally, the Cal.4th negative

bar sits modestly above its Cal.3d counterpart, while the Cal.4th positive bar sits modestly below its Cal.3d counterpart; together those movements generate the crossover rather than a spike in just one category.

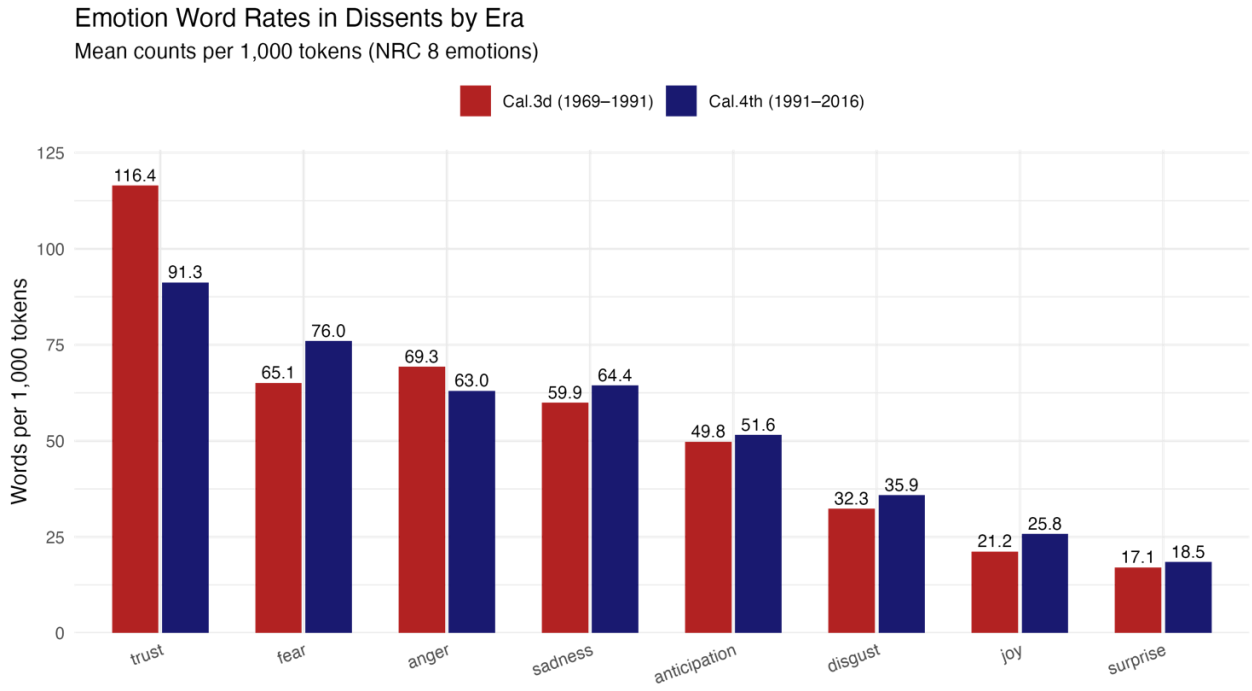

Figure 6.2: Emotion Word Rates in Dissents by Era

Having established a balanced but real shift in overall valence, the next question is what moved inside that valence. Figure 6.2 answers by disaggregating the NRC lexicon into its eight emotion families—trust, anticipation, joy, surprise (typically positive-coded) and fear, anger, sadness, disgust (typically negative-coded)—and plotting their opinion-level rates by era.

The tallest bar in Cal.3d is *trust*; in Cal.4th, *trust* drops noticeably. Cal.4th, by contrast, shows higher *fear*, *sadness*, and *disgust* rates. Those three emotions tend to accompany cautionary or harm-focused arguments (risk, error, institutional cost), so their rise helps explain why Cal.4th's negative bar exceeds its positive bar in Figure 6.1. The *anger* category is relatively close across

eras, with a slight advantage to Cal.3d, consistent with pointed disagreement appearing in both periods. *Anticipation* is similar (marginally higher in Cal.4th), while *joy* and *surprise* remain low in both eras and do not drive overall polarity.

Put together, the two figures tell a coherent story about tone. Earlier dissents more often speak in the register of assurance and confidence (*trust* leading, positive > negative). Later dissents more often lean on language of concern and cost (*fear/sadness/disgust* leading, negative > positive). The magnitudes are modest but steady across the corpus, suggesting a real shift in how disagreement is voiced.

6.2 Emotion Dynamics Over Time

Carrying the polarity measure from section 6.1 onto a timeline, Figure 6.3 traces year-by-year shifts in dissent tone. Each gray point is the annual mean of document-level polarity scores (negative − positive words per 1,000 tokens), with point size proportional to the number of dissenting opinions that year. A thin gray line connects yearly means. The midnight-blue curve is a LOESS smooth showing longer-run movement, and the light blue ribbon is its 95% confidence band. The dashed vertical line at 1991 marks the Cal.3d/Cal.4th boundary to contextualize the era shift. The polarity series centers on a zero line: values above zero indicate years when, on average, dissents used more negative than positive words per 1,000 tokens; values below zero indicate the reverse.

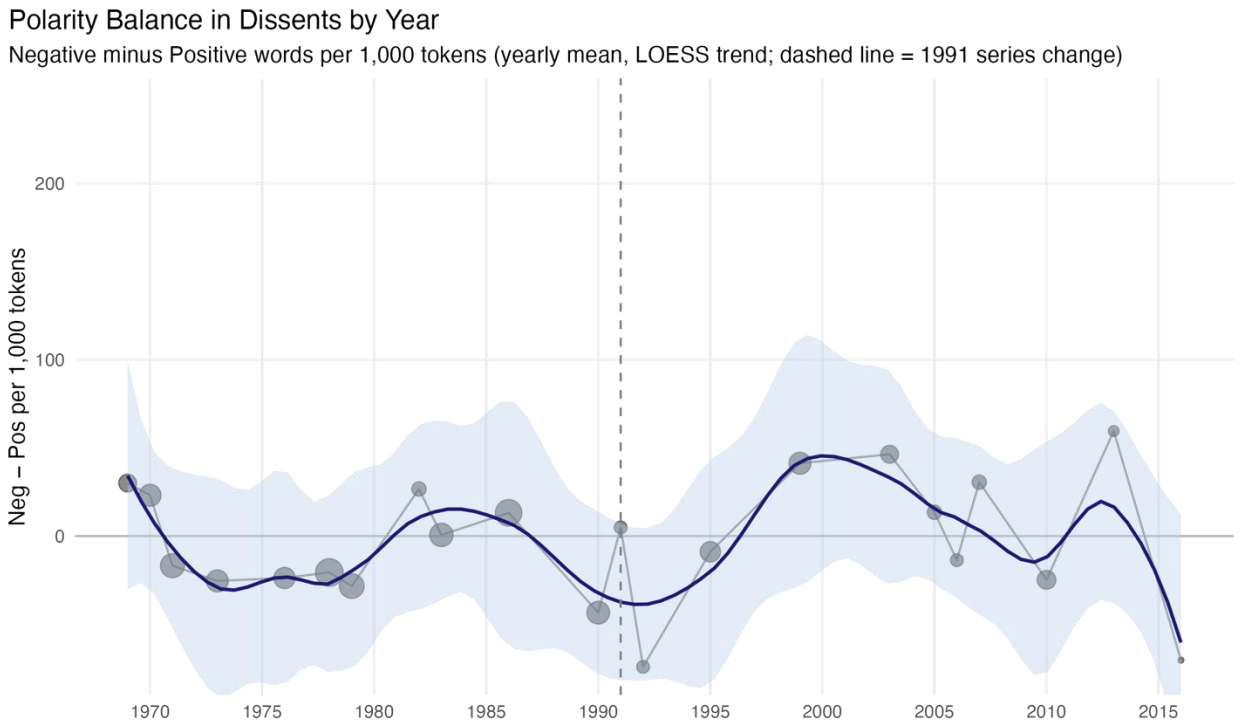

Figure 6.3: Polarity Balance in Dissents by Year

The series starts slightly above zero in 1969, dips below zero in the early 1970s (more positive than negative), and stays modestly below zero through the late 1970s. It moves back toward neutral in the early–mid 1980s with a few brief moves above zero, then drops sharply around 1991—a volatile trough indicated by a small point and wide band (few dissents, high uncertainty). From the late 1990s into the early 2000s it climbs decisively above zero, peaking around 2000–2002 (more negative than positive). The mid-2000s drift toward neutral, a below-zero dip appears around 2008–2010, and the early 2010s briefly rebound above zero before tapering toward or below zero at the end of the series, where uncertainty widens as yearly counts thin. In short: the early years lean slightly more positive, the turn of the millennium is clearly more negative, and the final years oscillate with greater uncertainty.

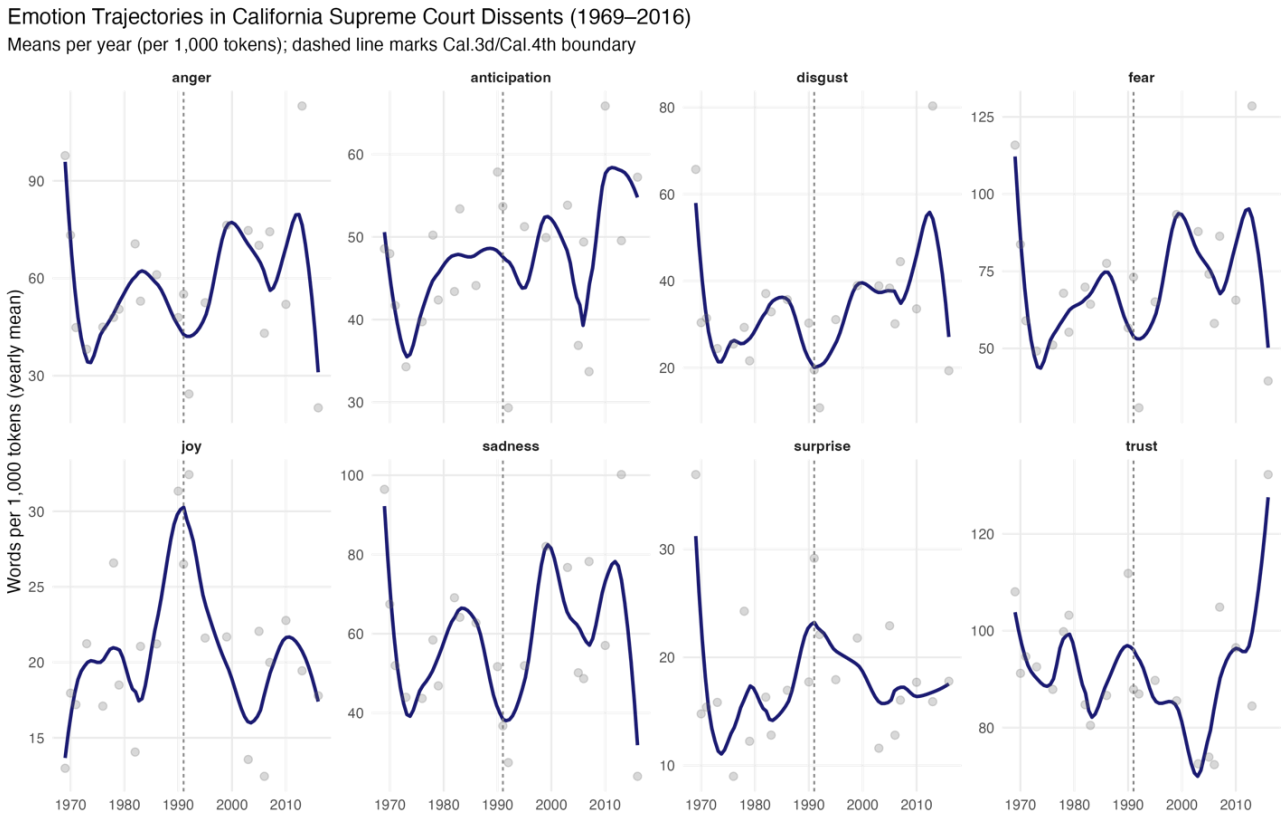

Figure 6.4: Emotion Trajectories in Dissent (1969-2016)

Figure 6.4 decomposes the annual signal into eight NRC emotions. For each dissent, tokens were mapped to a single emotion and scaled to counts per 1,000 words; those document rates were then averaged within year. In each panel, gray dots mark the yearly mean (larger dots = more dissents), a thin gray line links adjacent years, a LOESS curve traces the longer-run movement, and a dashed vertical at 1991 aligns the Cal.3d/Cal.4th eras. Several patterns align tightly with the polarity curve: Trust declines gradually from the 1970s toward a trough around 2000, then rises sharply after 2010. This arc helps explain both the net-negative bulge at the millennium and the brief return toward net-positive in the early 2010s. Fear, disgust, and sadness all exhibit minima near the early 1990s and pronounced increases into the late 1990s/early 2000s, peaking close to the polarity maximum. Those rises are the main drivers of the net-negative period around 2000. Anger shows a similar

trough around 1991 and a climb into the 2000s, though with more mid-decade variability than fear/disgust/sadness. Its increase contributes to, but does not solely determine, the 2000–2002 peak. Anticipation is broadly stable, with a mild rise around 1999–2000, a dip near 2010, and partial recovery thereafter. Joy remains low overall, with a small spike in the early 1990s and a modest bump around 2010; its scale is too small to move polarity materially. Surprise is the flattest and lowest of the eight, inching up in the 1990s and easing back afterward; it plays only a minor role in polarity shifts.

Together, the polarity curve and the emotion facets tell the same story: the early years lean slightly more positive (polarity < 0), a marked above-zero stretch around the millennium reflects more negative than positive language driven chiefly by rises in fear, disgust, and sadness alongside a trough in trust, and the 2010s show partial re-balancing as trust rebounds and those negative emotions ease. In other words, the millennium peak isn't a blanket surge in "negativity" but the coordinated lift of specific cautionary registers coupled with a contemporaneous decline in trust.

CHAPTER 7

Thematic Evolution

7.1 Dominant Topics in Dissents

This section presents an inventory of the principal themes that recur in California Supreme Court dissents and contrasts how their emphasis shifts from the Cal.3d series (1969–1991) to the Cal.4th series (1991–2016). The figures report, for each era, the highest-probability words for each learned topic; bars represent the topic–word probabilities and therefore indicate the most characteristic vocabulary of that topic.

Topics were estimated with Latent Dirichlet Allocation [JWF18] fitted separately to the Cal.3d and Cal.4th corpora, each with K = 7 topics. The same dissent-only, cleaned vocabulary from Chapter 6 was used, with trimming of rare and overly ubiquitous tokens and a small legal stoplist. Estimation employed collapsed Gibbs sampling with mildly sparse Dirichlet priors (α = 0.1) on document–topic mixtures and η = 0.05 on topic–word distributions). The sampler assigns the ith token (word type v) in document d to topic k with probability proportional to

$$p(z_{di} = k \mid z_{-di}, w, \alpha, \eta) \propto \left(n_{dk}^{-i} + \alpha_k\right) \times \frac{n_{kv}^{-i} + \eta_v}{n_{k\cdot}^{-i} + \sum_v \eta_v} \tag{7.1}$$

where n_{dk}^{-i} counts tokens in d currently assigned to k , n_{kv}^{-i} counts word v in topic k, and n_k^{-i} is the topic total (all excluding the i_{th} token). Topics are unlabeled by the model; descriptive names are assigned post hoc from the displayed top words.

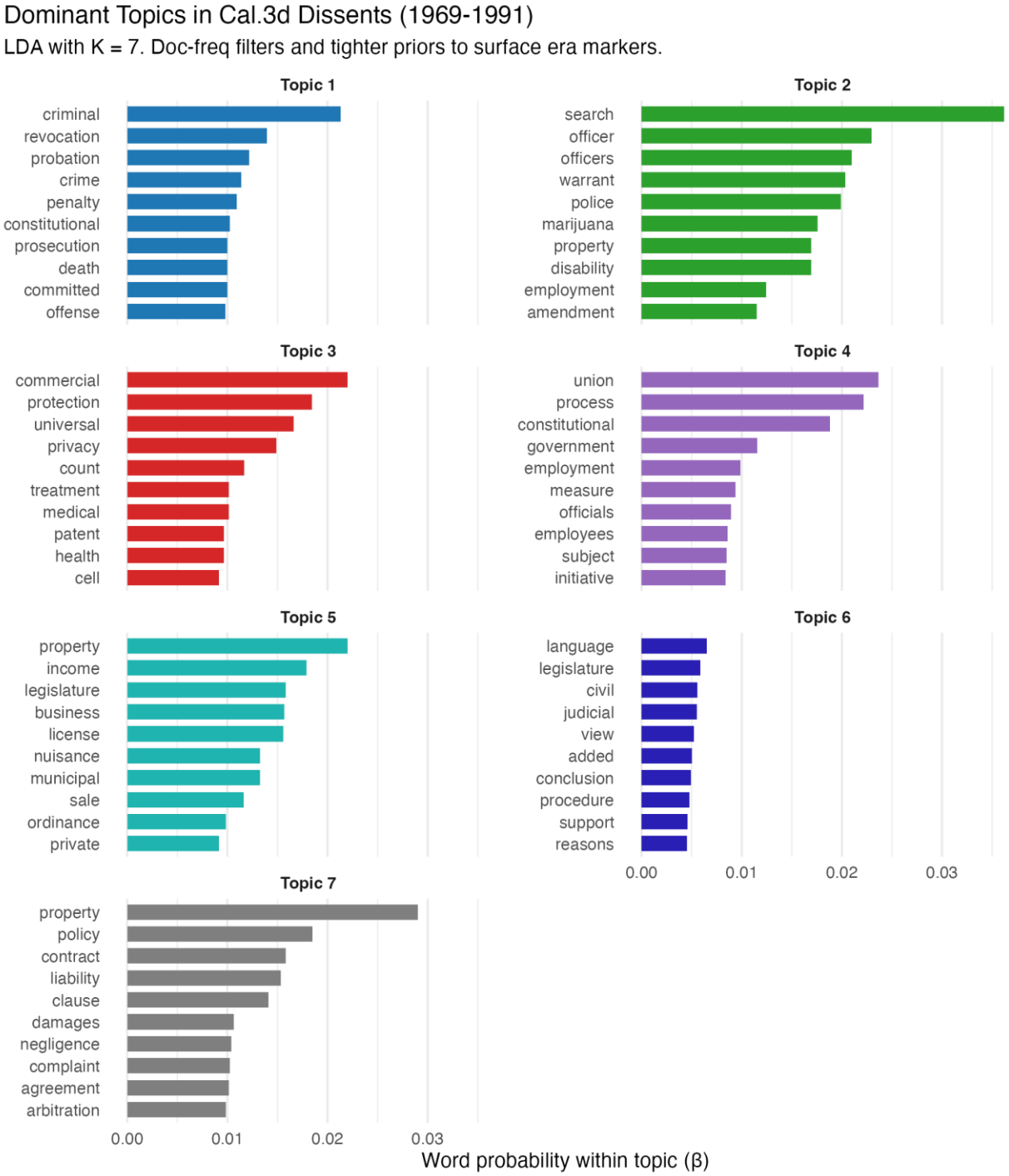

Figure 7.1: Dominant Topics in Cal.3d Dissents (1969-1991)

Shown in Figure 7.1 (Cal.3d), Topic 1 concentrates on charging, probation, and sentencing consequences—*criminal, revocation, probation, penalty, conviction, prosecution*. Topic 2 isolates search-and-seizure and police-conduct vocabulary—*search, officer/officers, warrant, police, marijuana, privacy*. Topic 3 gathers a consumer/medical-privacy cluster—*commercial, protection, medical, patent, health, privacy*. Topic 4 captures workplace and public-administration disputes—*union, employment, officials, employees, measure, initiative*. Topic 5 tracks municipal regulation and property governance—*property, income, license, nuisance, municipal, ordinance*. Topic 6 collects judicial and statutory phrasing—*language, legislature, civil, judicial, procedure, reasons*—a general interpretive register that recurs across opinions rather than a single doctrinal domain. Topic 7 returns to private-law liability and contracting—*property, contract, liability, clause, damages, negligence, arbitration*. Taken together, the Cal.3d facets show dissent rhetoric anchored in adjudicative safeguards (Topics 1–2) and civil governance/remedies (Topics 4–5, 7), with a cross-cutting procedural register (Topic 6).

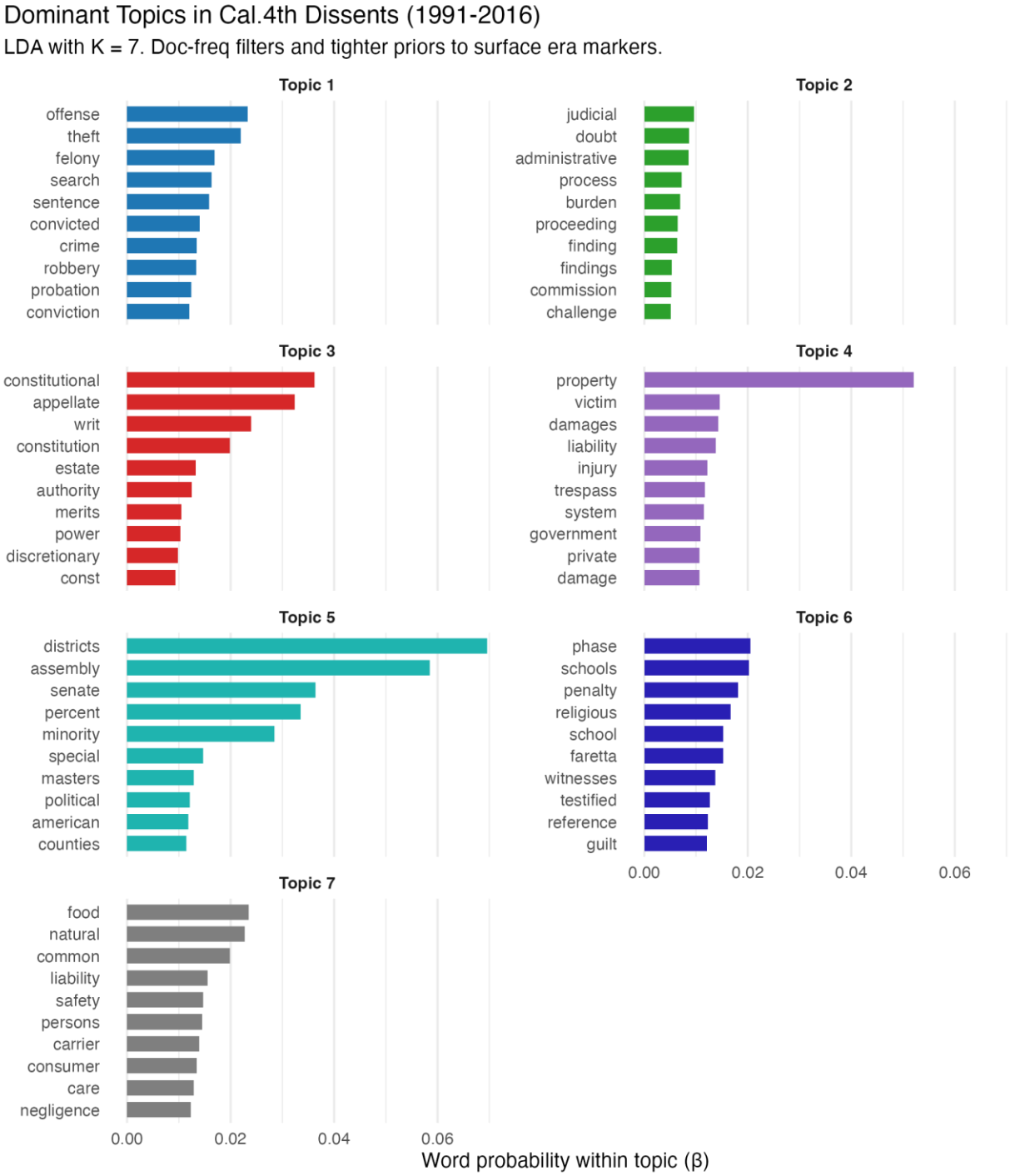

Figure 7.2: Dominant Topics in Cal.4th Dissents (1991-2016)

Described in Figure 7.2 (Cal.4th), Topic 1 centers on charging and sentencing—*offense, theft, felony, sentence, convicted, probation*. Topic 2 captures agency and commission proceedings and their review—*judicial, administrative, process, proceeding, findings/finding, commission, burden, challenge*. Topic 3 collects collateral-review and appellate mechanics—*constitutional, appellate, writ, merits, discretionary*. Topic 4 is the torts/property bucket—*property, victim, damages, liability, injury, trespass*. Topic 5 is the redistricting-and-equality theme—*districts, assembly, senate, percent, minority, special masters, counties*. Topic 6 isolates penalty-phase and trial-rights vocabulary—*phase, penalty, witnesses, Faretta, representation*—distinct from the broader charging topic. Topic 7 summarizes consumer-protection and product-safety disputes—*food, natural, safety, consumer, carrier, negligence*. In combination, the Cal.4th facets highlight structured sentencing (Topics 1 and 6), a specialized redistricting/equality domain (Topic 5), and modern consumer-protection alongside stable civil-liability and review topics (Topics 3–4, 7).

Read across the two figures, the earlier era gives greater prominence to search/seizure, probation/revocation, and municipal or workplace governance, while the later era adds distinctive statutory architectures—penalty-phase practice and redistricting/equality—without displacing enduring property/torts and review mechanics. Because panels display within-topic word probabilities, these contrasts reflect shifts in rhetorical focus rather than differences in case counts.

7.2 Era-Specific Theme Transitions

This section quantifies how the thematic composition of dissents reallocated from the Cal.3d series (1969–1991) to the Cal.4th series (1991–2016). The outcome plotted in Figure 7.3 is the average document-level share of each theme within an opinion: for each dissent, I aggregate its LDA mixture weights to a set of cross-era categories and then average those shares across all dissents

in the era. The statistic therefore captures reallocation of rhetorical space within opinions rather than changes in case counts.

To make the eras commensurable, the era-specific topics are consolidated into five substantive categories that both periods actually populate: criminal justice; administration & agencies; private law & markets; public governance & elections; and rights & liberties. One broad Cal.3d topic—procedural/interpretive language that recurs across many opinions—would otherwise swamp a single category. To avoid that distortion, I fractionally assign it 60% to administration & agencies (process-heavy phrasing) and 40% to rights & liberties (constitutional framing). Varying this split between 50/50 and 70/30 leaves the qualitative results unchanged, and medians yield the same ordering of changes.

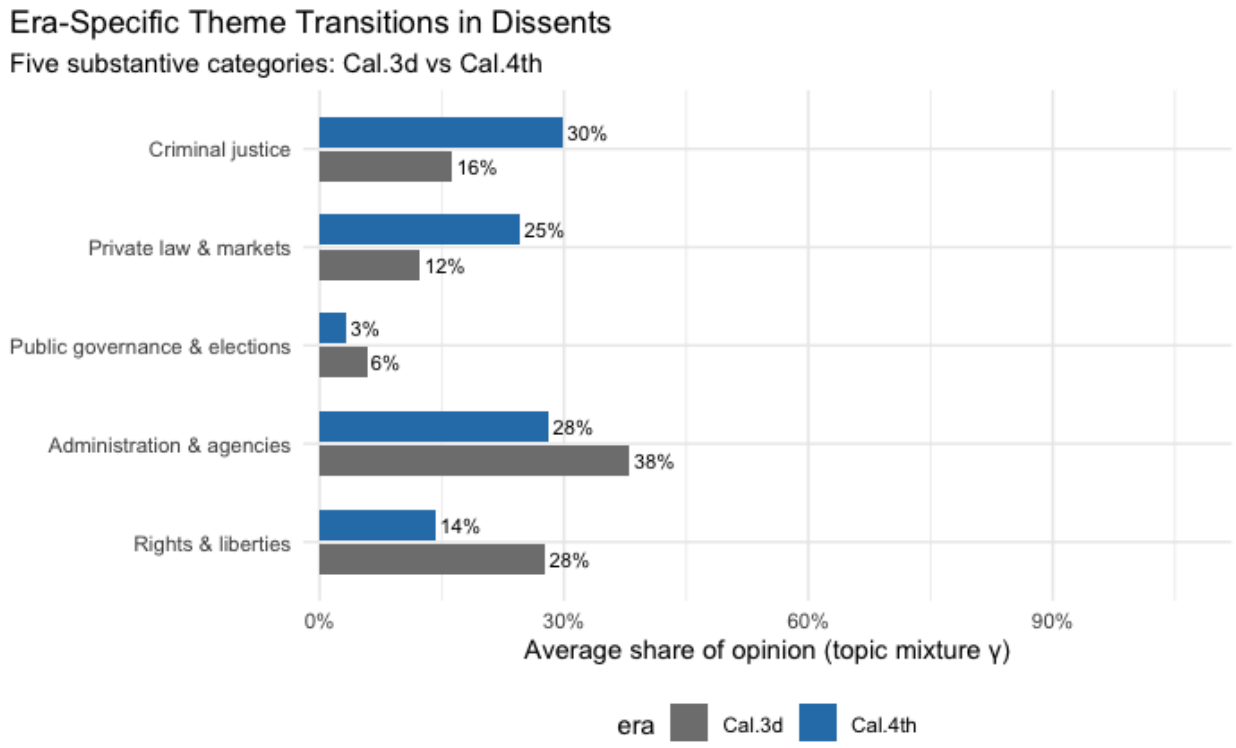

Figure 7.2: Era-Specific Theme Transitions in Dissents

Figure 7.3 shows a marked rise in criminal-justice share in Cal.4th, consistent with the era's emphasis on charging/sentencing and penalty-phase practice. Private law & markets also expands, reflecting steady attention to torts, property, contracts, and consumer/product-safety disputes. By contrast, administration & agencies occupies a larger share in Cal.3d once generic procedural verbiage is not allowed to dominate, indicating that employment/union governance and agency-process disputes had a comparatively bigger footprint in earlier dissents. Public governance & elections remains modest in both eras—municipal/initiative issues in Cal.3d and redistricting/equality in Cal.4th are visible but episodic—while rights & liberties persists across periods yet claims a smaller slice in Cal.4th after accounting for the growth of criminal-justice and private-law content.

CHAPTER 8

Conclusion

In this study, we have investigated the linguistic, rhetorical, and thematic evolution of dissenting opinions in the California Supreme Court, specifically comparing the Third Series (Cal.3d, 1969-1991) and the Fourth Series (Cal.4th, 1991-2016). By bridging legal history with computational linguistics, this research mapped the court's transition from an era of "independent state grounds" and judicial activism to one of stabilization and restraint. This chapter synthesizes the empirical evidence gathered through NLP techniques, articulates the study's contributions to legal scholarship, and acknowledges the methodological limitations while proposing avenues for future inquiry.

8.1 Research Summary

The methodological foundation of this study was established through the rigorous curation of a comprehensive corpus sourced from the Harvard Caselaw Access Project (CAP). Spanning nearly five decades of judicial history, the dataset encompassed the full text of dissenting opinions from both the Third (1969–1991) and Fourth (1991–2016) Series of the California Official Reports. To transform raw judicial output into analyzable data, we engineered a specialized preprocessing pipeline designed to navigate the syntactic complexities of legal text, stripping away procedural "boilerplate" and embedded citations to isolate the rhetorical core of the minority voice.

Building upon this cleaned corpus, the analytical framework integrated three distinct computational techniques to capture the multi-dimensional nature of dissent. We employed Term Frequency-Inverse Document Frequency (TF-IDF) to identify the unique lexical signatures of each era; the NRC Emotion Lexicon to quantify the subtle shifts in judicial sentiment; and Latent Dirichlet Allocation (LDA) to uncover latent thematic structures. By assessing these models against the historical watershed of 1991, we were able to measure rhetorical evolution using precise metrics such as weighted log-odds and topic probability distributions. This multi-method approach demonstrated that computational tools can effectively replicate and quantify complex historical narratives, transforming the qualitative concept of "judicial restraint" into measurable linguistic data.

8.2 Contribution and Significance to the Field

The primary contribution of this research is the empirical validation of the California Supreme Court's historical narrative, offering quantitative proof that the shift from the Bird Court to the Lucas and George Courts fundamentally altered the language of the law. While legal scholars have traditionally relied on doctrinal analysis to describe the court's retreat from "independent state grounds" and "judicial activism," this study demonstrates that this ideological shift is measurable in the micro-patterns of text. The data confirms that the 1986 retention election did not merely change case outcomes but reshaped the court's institutional identity, forcing the minority voice to abandon the confident lexicon of constitutional remedies in favor of a narrower, statute-bound vocabulary. This provides the legal community with a new lens for understanding "New Judicial Federalism," showing that judicial philosophies are expressed as much through emotional tone and vocabulary choice as through legal holding.

Furthermore, this work advances theoretical understanding of judicial polarization by establishing measurable patterns of rhetorical dissent. By isolating the emotional architecture of minority opinions, the analysis reveals that dissent serves a dynamic function that evolves in response to external political pressures. The identification of rising negative sentiment, specifically the increase in fear and disgust in the Cal.4th era, suggests that in periods of "judicial restraint," dissent becomes a more charged instrument, utilizing the rhetoric of alarm to contest the majority's deference to the electorate. This enriches the academic discourse on persuasive strategies within state courts, highlighting how justices use language to signal the intensity of their disagreement beyond the legal merits of a case.

From a methodological standpoint, this thesis delivers a replicable framework for comparative legal analysis that can be applied across jurisdictions. By successfully adapting NLP techniques to the unique constraints of legal texts, such as handling embedded citations, parsing formal structures, and filtering specialized terminology, this study solves key technical challenges that often hinder computational legal studies. It creates an analytical blueprint for future scholarships to move beyond manual case studies, enabling large-scale, data-driven explorations of how dissenting opinions function as instruments of legal development across different state supreme courts.

8.3 Limitation and Future Work

While this study offers significant insights, it is subject to certain limitations that suggest directions for future research. One primary constraint is the reliance on dictionary-based sentiment analysis. The use of the NRC Emotion Lexicon, while effective for identifying broad trends, relies on fixed word associations that may miss context-specific sarcasm or the unique legal usage of certain

emotional terms. Future work could employ context-aware models, such as transformer-based embeddings, to capture more subtle tonal nuances that a "bag-of-words" approach might overlook.

Additionally, this analysis focused exclusively on dissenting opinions to isolate the minority voice. However, dissents exist in conversation with a majority opinion. Future studies could analyze the divergence between majority and dissenting texts within the same case to measure the "semantic distance" between justices. This would provide a more granular view of ideological polarization and the specific dynamics of the "dialogue" between the court's factions. Finally, because this framework is currently limited to California, expanding the methodology to other state supreme courts would allow for comparative analysis. Such an expansion could determine if the shift toward statutory rhetoric and negative sentiment is unique to California's post-1986 political trauma or indicative of a broader national trend in American jurisprudence.

REFERENCES

[And72] *People v. Anderson*, 6 Cal. 3d 628 (1972).

[BDI18] Emily M. Bender, Leon Derczynski, and Pierre Isabelle. 2018. *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, Santa Fe, New Mexico, USA.

[BMB24] Bifari, E., Basbrain, A., Mirza, R., Bafail, A., Albaradei, S., & Alhalabi, W. (2024). Text mining and machine learning for crime classification: using unstructured narrative court documents in police academic. *Cogent Engineering*, *11*(1). https://doi.org/10.1080/23311916.2024.2359850

[Cap] Harvard Law School Library Innovation Lab. *Caselaw Access Project*. https://case.law

[Cha23] Chai, Christine P. "Comparison of Text Preprocessing Methods." *Natural Language Engineering* 29, no. 3 (2023): 509–53. https://doi.org/10.1017/S1351324922000213.

[DTA24] Krish Didwania, Durga Toshniwal & Amit Agarwal. "Unveiling Themes in Judicial Proceedings: A Cross-Country Study Using Topic Modeling on Legal Documents from India and the UK." *arXiv preprint*, arXiv:2406.00040v1 (2024). https://arxiv.org/abs/2406.00040

[ELP11] Lee Epstein, William M. Landes & Richard A. Posner. “Why (and When) Judges Dissent: A Theoretical and Empirical Analysis.” *Journal of Legal Analysis*, Vol. 3, No. 1 (2011): 101–137.

[GCC20] Mark P. Gergen, David A. Carrillo, Benjamin Minhao Chen & Kevin M. Quinn. *Partisan Voting on the California Supreme Court*. Southern California Law Review, Vol. 93, No. 4, 2020.

[JWF18] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li & Liang Zhao. “Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey.” *arXiv preprint*, arXiv:1711.04305v2 [cs.IR] (2018). https://doi.org/10.48550/arXiv.1711.04305

[KK24] Kucherenko, Yehor & Kulakovska, Inessa. (2024). Analysis of methods and algorithms for processing unstructured text data based on JSON technology. Technology audit and production reserves. 3. 10-18. 10.15587/2706-5448.2024.306435.

[Lan85] *In re Lance W.*, 37 Cal. 3d 873 (1985).

[LWZ18] Q. Liu, J. Wang, D. Zhang, Y. Yang and N. Wang, "Text Features Extraction based on TF-IDF Associating Semantic," 2018 IEEE 4th International Conference on Computer and Communications (ICCC), Chengdu, China, 2018, pp. 2338-2343, doi: 10.1109/CompComm.2018.8780663.

[Mar08] *In re Marriage Cases*, 43 Cal. 4th 757 (2008).

[MCQ08] Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis*, *16*(4), 372–403. http://www.jstor.org/stable/25791946

[Mil16] Kenneth P. Miller. "The California Supreme Court and the Popular Will." *Chapman Law Review*, Vol. 19, 2016.

[MT13] Saif M. Mohammad & Peter D. Turney. *NRC Word-Emotion Association Lexicon*. National Research Council Canada, 2013.

[Nay24] Nay, John, Natural Language Processing and Machine Learning for Law and Policy Texts (April 7, 2018). Nay, J. (2021) "Natural Language Processing for Legal Texts." In D. M. Katz, R. Dolin & M. Bommarito (Eds.), Legal Informatics. Cambridge University Press., https://ssrn.com/abstract=3438276 or http://dx.doi.org/10.2139/ssrn.3438276

[Opp10] David B. Oppenheimer & Allan Brotsky (Eds.). *The Great Dissents of the "Lone Dissenter": Justice Jesse W. Carter's Twenty Tumultuous Years on the California Supreme Court*. Carolina Academic Press, 2010.

[Sch16] Harry N. Scheiber (Ed.). *Constitutional Governance and Judicial Power: The History of the California Supreme Court*. Berkeley Public Policy Press, 2016.

[TWL02] Chade-Meng Tan, Yuan-Fang Wang & Chan-Do Lee. "The Use of Bigrams to Enhance Text Categorization." *Information Processing & Management*, Vol. 38, No. 4 (2002): 529–546. https://doi.org/10.1016/S0306-4573(01)00045-0