"Undocumented Immigrants" != "Illegal Aliens": Decomposing the Conceptual and Narrative Landscapes of Partisan Immigration Terms

Yejin Cho[†], Gabriella Chronis[†], Nitin Sudarsanam[◇], Kevin Barcenas-Martinez[†], Katrin Erk[♣]

[†] The University of Texas at Austin, Austin, TX

[◇] Brown University, Providence, RI [♣] University of Massachusetts Amherst, Amherst, MA

{ycho, gabriellachronis}@utexas.edu, kevinbarcenas2022@gmail.com

nitin_sudarsanam@brown.edu, kerk@umass.edu

Abstract

Do politically charged terms with similar referents, like *undocumented immigrants* (UI) *illegal aliens* (IA), differ only in who uses them, or also in what they mean? We investigate usage patterns by projecting contextual embeddings into interpretable psycholinguistic feature space, and extracting narrative scenes with LLMs. We find that in partisan news, the term IA appears in contexts emphasizing causation and fear. UI appears in contexts emphasizing consequences experienced and shared humanity. Scene abstraction reveals parallel patterns: IA is embedded in narratives of criminality and threat, UI in narratives of vulnerability and governance. Beyond indexing speaker identity, these terms impart different construals on migrants: as *agents* of harm versus *patients* of circumstance. This dual-track methodology adds new tools to the growing body of computational approaches for understanding the conceptual framing of political discourse.[1]

1 Introduction

Political discourse often employs different terms for the same referent, with word choice reflecting ideological stance. When speakers choose between "undocumented immigrants" and "illegal aliens," they are signaling political identity. But what meaningful work does lexical choice accomplish beyond social indexicality? We investigate whether these partisan alternatives differ systematically in meaning, and if so, how.

Prior work has established that partisan vocabulary differences are detectable and predictable. Webson et al. (2020) used word embeddings to construct separate subspaces for denotational and connotational meaning, demonstrating that words with similar referents but different connotations, including "undocumented immigrants" and "illegal aliens," occupy similar positions in denotational space but diverge in connotational space, with this divergence tracking political orientation. Their approach successfully groups together words with strong left-leaning or right-leaning associations, telling us for whom these words differ, but not what makes them different semantically.

We extend this line of work by asking: what are the conceptual properties and narrative landscapes that distinguish partisan vocabulary? To answer this question, we employ a dual-track approach combining two complementary methods. First, we project contextual embeddings into an interpretable semantic feature space derived from psycholinguistic norms (Binder et al., 2016), allowing us to identify what properties each phrase activates, such as Human, Bright, Fear, or Social. Second, we use LLM-based scene abstraction to extract the narrative contexts (events, participants, and emotional valences) in which each phrase is embedded. These tracks answer complement and reinforce each other, together providing a richer picture than either alone.

Using a corpus of partisan news sources (Kiesel et al., 2019), we find that "undocumented immigrants" and "illegal aliens" activate systematically different semantic features. "Undocumented immigrants" is associated with features emphasizing consequences experienced, human qualities, and suffering (Consequential, Human, Needs). While "illegal aliens" is associated with features emphasizing agency and embodied experience, qualitative analysis shows that the phrase appears in contexts where immigrants cause harm. The independent scene abstraction analysis confirms this picture, revealing that the phrase "illegal aliens" is used to tell narratives of criminality and security threat, while "undocumented immigrants" is used to tell stories of reform and humanitarian concern. These results suggest that partisan word choice is not merely a badge of identity but both shapes and participates

[1] Code and analyses available at https://github.com/gchronis/partisan_news_features

*Proceedings of the 15th Joint Conference on Lexical and Computational Semantics (*SEM 2026)*, pages 348–365
July 3, 2026 ©2026 Association for Computational Linguistics

in distinct construals of its referent.

There is a wealth of research, both computational and otherwise, about the criminal/victim framing of immigration discourse. The main contribution of this paper is 1) applying two recent methods for computational lexical semantics to the analysis of social meaning, which 2) strengthen existing evidence about immigration framing and 3) provide additional insight about the relationship between narrative framing and agency.

As much as recent trends in NLP appear to depart from classical distributional semantics, they continue to rely on the foundational insight that meaning can be understood through relationships in meaning space. Large language models (LLMs) are also large culture models, encoding not just linguistic regularities but the social and ideological structures reflected in their training data. Understanding how these models represent politically charged concepts is thus both a scientific question about meaning and a practical question about the cultural knowledge these systems absorb and reproduce.

2 Related Work

2.1 Computational Analysis of Political Language

Partisan language differences in political communication are well documented (Gentzgow and Shapiro, 2010; Beaver and Stanley, 2023). Classification-based approaches can identify polarization (Peterson and Spirling, 2018), estimate ideological placement (Rheault and Cochrane, 2020), and detect partisan vocabulary (Webson et al., 2020). However, knowing that a phrase leans left or right does not reveal the conceptual frames within which these lexical items function.

Frame analysis offers a popular descriptive approach for immigration discourse. George Lakoff contrasted the *Illegal* and *Undocumented* frames, arguing that each activates distinct conceptual structures (Lakoff and Ferguson, 2006). Discourse-analytic work has uncovered recurring metaphors of migrants as *water*, *invaders*, *unwanted guests*, and *resources*, *inter alia* (Khosravinik, 2010; Taylor, 2021; Zawadzka-Paluektau, 2023; Cisneros, 2008; Gabrielatos and Baker, 2008).

Notably, Zawadzka-Paluektau (2023) find that objectifying metaphors (e.g., migrants as *resources*) are lexicalized more frequently than overtly hostile ones (e.g., *invaders*). Our approach complements work on lexicalized metaphor by uncovering semantic frames that are active even in the absence of overt metaphorical language.

Computational approaches to frame analysis have sought to scale these insights through annotation (Mohler et al., 2016), interactive refinement (Pacheco et al., 2023), and transformer-based analysis of emotional connotation (Hosseini and Staab, 2024). However, little work examines how partisan term choice correlates with systematic differences in semantic framing.

2.2 Distributional Semantics and Immigration

Word embeddings have become a prominent tool for studying social dimensions of meaning, from gender bias (Bolukbasi et al., 2016) to cultural associations along class, race, and gender (Kozlowski et al., 2019). Distributional methods have also been applied to political concepts: Dahlberg et al. (2020) study cross-cultural variation in the meaning of *democracy*. However, dimensions of analysis are typically handpicked. Recent work demonstrates that contextual embeddings can be mapped to psycholinguistic features, enabling interpretable analysis of word meaning in context (Turton et al., 2021; Chersoni et al., 2021; Apidianaki and Soler, 2021; Proietti et al., 2022; Chronis et al., 2023; Ranganathan et al., 2025). We apply this method to the study of social and political meaning.

2.3 Contextualizing Lexical Meaning via Situations and Stories

Frame Semantics argues that lexical meaning is rooted in systems of background knowledge about typical situations, including cultural and affective influences (Fillmore, 1976). Early computational work on extracting knowledge about such schemas focused on learning them through participant co-occurrence across events (Chambers and Jurafsky, 2008, 2009). More recently, LLMs have facilitated more flexible modeling. Frameworks like COMET-ATOMIC (Sap et al., 2019; Hwang et al., 2021) and DREAM (Gu et al., 2022) elicit situational dimensions such as emotional consequences and preconditions in order to formalize the commonsense knowledge required to reconstruct a narrative. This process of externalizing situational features provides a window into the "story space" (Erk and Chronis, 2022) encoded within contextualized embeddings.

However, effectively navigating this "story space" requires a framework capable of externalizing scenes into explicit and interpretable representations. We use the Scene Abstraction framework (Cho and Erk, 2026), which provides a method for externalizing latent scene structures into explicit, human-readable representations. Our approach maps natural language into a human-designed scene schema. We use LLMs to populate the dimensions of this schema, generating a structured, interpretable representation of the scene. By grounding highly polarized terms in these Scene Abstractions, we uncover the systematic correspondence between lexical choice and the underlying scene structures they frequently convey. Like semantic frames, Scene Abstractions are tied to a particular lexical unit; however, they target broader cultural and situational contexts than FrameNet (Baker, 2014).

3 Data and Methods

3.1 Hyperpartisan News Dataset

We utilize the Hyperpartisan news dataset (Kiesel et al., 2019), which contains over 750,000 news articles labeled with publisher-level partisanship (far-left, center-left, center, center-right, and far-right). We parse raw article text into sentences, merging quotations back into their enclosing sentences to maintain context.

We then extract sentences containing the phrases *undocumented immigrants* and *illegal aliens* (hereafter UI and IA, respectively). In keeping with Webson et al. (2020)'s study of phrases with shared referents and different connotation, we focus on the full phrases rather than individual words. In total, we collected a corpus of 4,323 samples, comprising 3,221 instances of UI and 1,102 of IA.

3.2 Dual-Track Representation of Meaning

To analyze the partisan framing of migration terminology, we represent each sentence through a dual-track approach that captures both cognitive associations and narrative contexts. In contrast to lexicon-based tools like LIWC (Pennebaker et al., 2015; Tausczik and Pennebaker, 2010), which has been productively applied to immigration discourse (Card et al., 2022) and parliamentary polarization (Gennaro and Ash, 2022), our methods leverage the transformer models which are sensitive to contextual variation, word order, pragmatic effects, syntactic dependency, and other higher-order phenomena.

3.2.1 Interpretable Semantic Features

First, we analyze the semantic properties of the target words, as indicated by contextualized embeddings. The dimensions of language model embeddings are uninterpretable in and of themselves. We follow Chronis et al. (2023)'s method for creating interpretable semantic spaces. This method involves training a deep neural network to predict psycholinguistic feature norms for a word-in-context from contextual embeddings for that word. Binder features rate target words on a scale of 1-7 for 65 cognitive features whose existence is grounded in neuroscience research, in Sensory, Motor, Space, Time, Social, and Emotion domains. We use the pretrained model from the `semantic-features` project (Ranganathan et al., 2025) for mapping layer 7 RoBERTa-base embeddings to Binder feature norms.

The Binder feature set was designed to be a comprehensive semantic breakdown. Though any feature set will offer its own lens on meaning rather than a universal decomposition, the features span experiential and emotional dimensions, and these dimensions end up organizing words in the training data according to cognitive affordances (Binder et al., 2016). Emotion, situated experience, and affordances are important aspects of social meaning and semantic frames. Unlike other feature norms like McRae et al. (2005) or Buchanan et al. (2019), each cue word in the dataset is rated on every dimension, and it spans concrete and abstract verbs, nouns, and adjectives. This contextual Binder embedding method has been applied to study lexical meaning variation in grammatical construal, cultural metapragmatics, and diachronic change (Chronis et al., 2023; Chronis, 2026), and is extended here to political connotation.

3.2.2 LLM Scene Abstraction

To capture situational structures that are latent to surface-level linguistic descriptions, we apply the **Scene Abstraction** framework (Cho and Erk, 2026). Given a sentence s and a target term w, the framework directs the LLM to generate a structured scene representation $\mathcal{S}(s, w)$ consisting of two complementary components:

The **Contextual Scene** $\mathcal{C}$ captures the broader situational landscape independent of w, consisting of: (1) *Events* (core actions in abstract, atomic form, following the COMET-ATOMIC style), (2) *Entities* (participants with roles, properties, and inferred emotions), and (3) *Setting* (spatio-temporal

INPUT SENTENCE

He urged a balanced approach, noting that those **undocumented immigrants** who commit serious crimes 'have no business in the United States.'

CONTEXTUAL SCENE $\mathcal{C}$

Events	PersonX urges a balanced approach PersonX notes the status of GroupY PersonX states GroupY should leave US
Entities	PersonX (*He*): Advocate Property: Reasoned Emotion: Concern GroupY (*undoc. immigrants*): Subjects Property: Vulnerable, Criminalized Emotion: Fear, Tension
Setting	Place/Time: Unspecified Atmosphere: Serious, Contentious

EXPRESSION PROFILE $\mathcal{E}$ *(undocumented immigrants)*

Eng. Events	PersonX discusses their status They involve in serious crimes
Gen. Props.	Political debate subjects Perceived as a mixed group with positive and negative associations
Evk. Emos.	Concern (legal status and societal impact) Tension (serious nature of crimes)

Figure 1: Simplified Scene Abstraction output for a sample sentence. In the Expression Profile ($\mathcal{E}$), *Eng. Events*, *Gen. Props.*, and *Evk. Emos.* denote Engaging Events, General Properties, and Evoked Emotions, respectively. See Appendix B.2 for full details.

and atmospheric backgrounds). The **Expression Profile** $\mathcal{E}$ provides the scene-grounded interpretation of w specifically, extracting: (1) *Engaged Events* involving w, (2) *Generalizable Properties* of w highlighted in context, and (3) *Evoked Emotions*, which is the affective impact of w within the scene.

Cho and Erk (2026) evaluate the Contextual Scenes against human judgments: They find that model predictions agree with human judgments on scene similarity in an odd-one-out evaluation; and they find that human raters assigned high average satisfaction values to scene descriptions generated by the model, and preferred them over commonsense inferences generated by ATOMIC (Hwang et al., 2021). Following Cho and Erk (2026), we implemented this model using `gpt-4o-mini` with a few-shot prompting strategy based on three manually curated in-context examples. The mapping into a pre-defined scene schema (see Figure 1; full output in Figure 4) enables systematic extraction of situational features while preserving the nuanced context necessary for sociolinguistic analysis.

In summary, we employ these two tracks to map each phrase choice (*illegal aliens* or *undocumented immigrants*) to both its latent cognitive properties and its narrative scaffolds, allowing us to analyze both *what* properties a phrase possesses (Section 3.2.1) and *how* those properties fit into narrative landscapes (Section 3.2.2).

4 Experiments

4.1 Feature Differences between Phrases

For each of the 65 Binder features, we performed linear regression predicting feature value from phrase identity ("undocumented immigrants" vs. "illegal aliens"), controlling for partisan label as a fixed effect. P-values were corrected for multiple comparisons using Bonferroni correction. After correction, 59 of 65 features showed significant differences between phrases ($p < 0.05$). Because the dataset imbalance could cause unstable coefficients and lower confidence for the minority class (IA), we report 95% confidence intervals and perform a downsampled analysis, balanced by partisanship, to ensure robustness. Regression coefficients between both sets of models were perfectly correlated (Pearson's $r = 1.0$).

Results revealed substantial differences in semantic feature activation across terms, as shown by the regression coefficients in Figure 2. Features most predictive of UI included Consequential, Needs, Human, LowerLimb, Complexity, UpperLimb (Motor domain), Speech, Short (Time domain), Angry, and Loud. In contrast, features most predictive of IA included Sound, Surprised, Self, Caused, Fearful, Music, Head, Body, Vision, and Weight. R-squared values ranged from near zero for non-significant features to 0.31 for Needs and 0.28 for Consequential, indicating that phrase identity alone explains a substantial portion of variance in feature activation for these dimensions.

Interpreting the Semantic Dimensions For the 15 most distinctive features, four of the authors looked for patterns in the ten highest- and lowest-scoring sentences. Annotators were blind to which feature they were examining and whether they were viewing high or low values. Interpretations are based on these annotations as well as a set of 'anchor words': the cue words from the model training

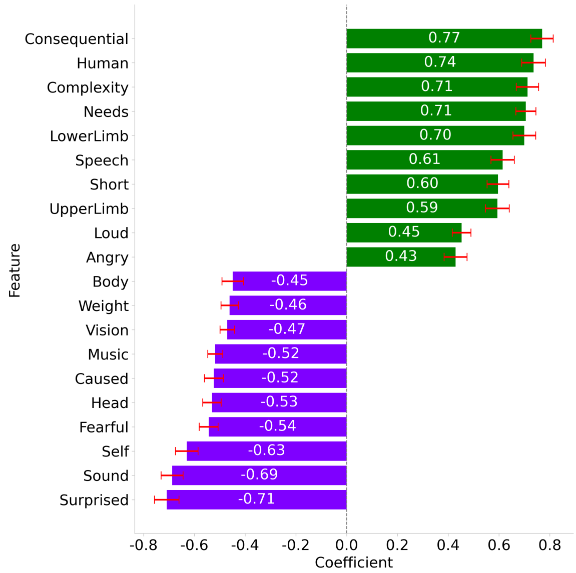

Figure 2: Binder Features with the 10 highest and lowest regression coefficients. Positive coefficients (Green) are higher for *undocumented immigrants*; negative (Purple) are higher for *illegal aliens*.

data that had the highest average human judgments. See Appendix A for the full set of example sentences and anchor words. For the top 5 distinctive features for each phrase, we report on those for which the authors all agreed on an easy interpretation. For Sound, Surprised, Self, and Needs, the authors did not immediately observe patterns relating high and low-valued sentences to the target feature. It may be that the statistical regularity in the representation of these features is related to less-visible features that are not instinctive for analysis.

Caused (with a high value predictive of IA) was the most clearly interpretable feature. Annotators independently converged on a striking pattern: sentences with high Caused values portrayed immigrants as agents of harm. These sentences described IA causing Arizona wildfires, being "responsible for seventy thousand sexual assaults," attempting to "choke the agent," and committing "193 homicide convictions, 426 sexual assault convictions, 1075 aggravated assault convictions." These examples are consistent with the anchor words *riot, destroyed, grievance, apology, trial, funeral*. In sentences with low Caused values, immigrants were acted upon: "their fates are decided," they are "threatened" by the Trump administration and "impacted" by policy decisions.

The inverse pattern appeared in the **Consequential** feature, where a high value predicted UI. High-Consequential sentences consistently portrayed immigrants as experiencing consequences such as "expedited removal" and "terror inducing raids," and might "lose their status." Annotators characterized these sentences as depicting immigrants as "victims."

Fearful sentences describe immigrants as presenting a dangerous threat. High-UpperLimb tokens (a Motor feature), more commonly UI, focus on deportation, whereas low-value tokens focused on money being spent on migrants, and characterize migrants as recipients or beneficiaries.

Human was more difficult to interpret. One annotator noted that High-Human sentences explicitly invoked humanity at the abstract level of human rights: migrants “have the right of due process," "enjoy a de facto amnesty" "universal human right to a nationality." **Self** was also difficult to interpret, but high valued-sentences discussed topics related to life experience: education, work, and voting, which is consistent with the anchor words *lived, worked, ate, slept, intellect*.

Discussion Together, the **Caused** and **Consequential** features reveal a fundamental asymmetry in how the two phrases construe agency. IA appears in contexts emphasizing actions, particularly harmful actions, of migrants, while UI more often discusses what is done *to* migrants. This pattern suggests that word choice correlates with, and perhaps cues, different narrative roles: perpetrator versus patient, threat versus victim.

Not all predictive features yielded clear interpretations. **Needs** and **Sound** ranked among the strongest predictors, yet annotators found these clusters difficult to characterize. As one annotator noted, "It's natural to look at the data in terms of events." An unblinded look revealed that high-Sound sentences frequently contained emphatic speech verbs (e.g. *claimed*, *huffed*) or placed the target phrase in direct quotation, which was not the case for any low-scoring sentences. Discourse structure and grammar can impact interpretation in ways that fall below the threshold of metapragmatic awareness (Silverstein, 2009). These regularities, though predictive, resist the content-focused intuitions annotators naturally bring to the task.

4.2 Scene Analysis: Discriminative Features

Methods To quantify situational information, we transformed the structured scene abstractions into binary feature vectors. The LLM outputs structured JSON; we extracted individual values from each field and converted them into binary tokens indicating presence or absence of a specific scene feature. Each token is constructed by concatenating the field name and its value with underscores. For example, from the scene output in Figure 1, the emotion *Fear* assigned to the target entity *undocumented immigrants* (`GroupY`) yields `target_entity_emotion_Fear=1`, while the property *Reasoned* assigned to the context entity `PersonX` yields `context_entity_property_Reasoned=1`. The prefix *target_* marks features of the focal term (IA or UI), and *context_* marks features of surrounding entities. Fields covered include all components of the scene schema: events, entity roles, properties, and emotions, setting attributes, engaged events, generalizable properties, and evoked emotions. After discarding features appearing fewer than five times ($min\_df = 5$), this procedure yielded a 949-dimensional feature vector per sentence. Event descriptions tend to yield longer, more specific tokens that often do not survive this filter, though recurring patterns such as `events_personx_signs_an_executive_order` are retained. The resulting 4,323 vectors were used to train a logistic regression model for predicting lexical choice (UI vs. IA), using a stratified 80:20 train-test split.

Results The logistic regression model achieved a classification accuracy of 82.2% (Macro F1: 0.745). Compared to the majority class baseline (Macro F1: 0.427), the model shows reliable performance across classes, validating its utility for our subsequent analysis of the scene features associated with IA and UI. The analysis of model coefficients in Table 1 shows distinct situational patterns associated with each phrase.

Features most predictive of UI: The term UI is strongly characterized by features that humanize the referents by highlighting their individual agency and internal psychological states. The most powerful predictor was the target's emotion of *Resilience* ($\beta = 3.71$), alongside *Hope* ($\beta = 1.93$). Furthermore, UI is frequently situated in contexts that highlight the referent's vulnerability (e.g., *Vulnerable population*, *Vulnerable status*) or proactive social roles, such as being *Applicants* ($\beta = 1.77$) or *Seeking legal recognition* ($\beta = 2.30$). These features suggest a narrative framing that emphasizes emotional empathy, respect for legality, and overcoming challenges.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[T] emotion: resilience`	3.71
`[P] subject to political challenges`	3.19
`[T] role: population`	2.72
`[T] role: vulnerable population`	2.41
`[T] prop: young`	2.37
`[T] prop: seeking legal recognition`	2.30
`[P] central to political debates`	2.06
`[T] property: lacking legal status`	2.06
`[T] property: vulnerable status`	2.03
`[T] emotion: hope`	1.93
Top Predictors for Illegal Aliens (IA)	
`[T] property: undocumented`	-4.44
`[T] property: assoc w/ criminality`	-2.51
`[P] subj to law enforcement actions`	-2.23
`[T] property: unauthorized`	-2.17
`[S] place: america`	-2.02
`[T] property: undocumented status`	-1.98
`[C] property: criminal`	-1.75
`[T] property: unlawful status`	-1.73
`[P] involved in policy discussions`	-1.66
`[T] property: mobile`	-1.66

Note: Prefixes are abbreviated as: **[T]**: Target Entity, **[C]**: Context Entity, **[P]**: Generalizable Properties (of the target terms), **[S]**: Setting.

Table 1: Top 10 Predictive Scene Features for UI and IA instances. Features are listed in descending order of $|\beta|$.

Features most predictive of IA: In stark contrast, the label IA is typically invoked in scenes saturated with themes of *criminality* and *Law enforcement actions* ($\beta = -2.23$). The attributes of Context Entities (surrounding actors) are also negative, described as *Criminal* ($\beta = -1.75$) or *Aggressive* ($\beta = -1.52$). A closer qualitative inspection reveals that these features often refer to adversarial figures such as smugglers, gang members, or perpetrators of violent crime, as well as political entities engaged in aggressive rhetoric. This last feature in particular suggests that the other entities mentioned in context with IA exhibit hostility towards them. This pattern suggests that the label IA exhibits a distributional preference for scenes of social and legal conflict.

Because some of the strongest predictors of lexical choice are features that are near synonyms of the target classes (e.g., unlawful status, undocumented), the distinction between scene descriptions for the two classes may seem stronger than it is. Nevertheless, this exploratory analysis surfaces themes in the scene features common to UI and IA, respectively.

4.3 Scene Analysis: Cluster-level Narratives

To analyze situational patterns holistically, we transformed the scene generated scene descriptions into vector representations using SBERT (`all-mpnet-base-v2`; Cho and Erk, 2026, found that SBERT over scene abstractions was a much better match for human scene similarity ratings than SBERT over raw sentences). We computed separate embeddings for the *events*, *target/context entities*, *place*, *time*, and *atmosphere* components of the Contextual Scene, as well as for the *engaged events*, *generalizable properties*, and *evoked emotions* components of the Expression Profile. All these embeddings were then concatenated to form a comprehensive scene vector. This vector captures the comprehensive narrative profile associated with a token.

To identify distinct narrative landscapes, we performed K-means clustering ($k = 5$) separately on the IA and UI scene sets (see Appendix B.3 for visualization).[2] To manage high dimensionality, we applied PCA to reduce the space to 50 principal components. The resulting clusters represent coherent scene profiles, providing a multi-dimensional view of how these terms are structurally embedded within different discourse landscapes.

4.3.1 N-gram Interpretation of Scene Clusters

To qualitatively interpret the emerging narrative landscapes, we constructed detailed scene profiles for each IA and UI cluster (see Tables 2 and 3). We build the profiles with the scene feature n-grams alongside raw sentence n-grams. While lexical indicators capture the specific surface-level language used, the scene-based n-grams provide a structural summary of the more abstract situational patterns that characterize each cluster. Thematic labels for each cluster were generated heuristically based on independent thematic descriptions by each author in combination with Gemini 3 Flash descriptions of each scene profile. These labels based on the independent interpretations each author, which agreed with one another and largely agreed with the LLM.

The analysis of these scene profiles reveals distinct narrative landscapes for each phrase:

[2] Because word meanings and conventionalized scenes are internally variable and not cleanly separated, different values for K yield different levels of granularity in meaning representation. There is no ideal K value; K=5 is significantly granular to capture some major story distinctions while remaining manageable.

Cluster	Scene Profiles
IA_0	**Theme: Administrative Criminalization** **Events:** identified, legal, individuals, aims, report, program **Entities:** [T] subjects, recipients (vulnerability); [C] authority (authoritative, determined) **Setting:** california, san diego; fiscal 2010; regulatory, formal **Surface N-grams:** convicted felons, vote, mental, billion, restrictions, incompetents, aimed, felons adjudicated **Features:** +: Happy, Cognition, Loud, Social, Fast –: Path, Pleasant, Shape, Body, Consequential
IA_1	**Theme: Political Weaponization** **Events:** amnesty, expresses, individuals, support, citizenship **Entities:** [T] subjects, recipients, policy (uncertainty); [C] speaker (influential, determined) **Setting:** interview setting; campaign, election; charged, contentious **Surface N-grams:** executive, reform, cruz, campaign, daca, executive amnesty, voting, democrats, comprehensive **Features:** +: Path, Motion, Pattern, Practice, Shape –: Happy, Harm, Cognition, Head, Pain
IA_2	**Theme: Security Threat** **Events:** individuals, identified, objecty, claims, expresses, border **Entities:** [T] subjects, victims, perpetrators (fear); [C] authority (concerned) **Setting:** border, texas; contemporary; tense, confrontational **Surface N-grams:** agent, ve, workers, thousands, assault, kill, got, communities, white, human **Features:** +: Disgusted, Attention, Arousal, Audition, Away –: Unpleasant, Loud, Speech, Social, Human
IA_3	**Theme: Media Outrage** **Events:** expresses, makes statement, claims, discusses, group **Entities:** [T] subject (distressed); [C] speaker, advocate (strong, hostile) **Setting:** media, public, studio; discourse; contentious, aggressive **Surface N-grams:** dobbs, saying, need, cruz, comments, executive, want, asked, bannon, bad **Features:** +: Harm, Toward, Unpleasant, Vision, Fearful –: Needs, Fast, Head, Fearful, Vision
IA_4	**Theme: Jurisdictional Conflict** **Events:** border, cities, sanctuary, deportation, authorities **Entities:** [T] subjects, recipients (potential fear); [C] authority, maker (power, determined) **Setting:** california, arizona; tense, charged, uncertain **Surface N-grams:** executive, case, department, arizona, 11, illegally, agents, agency, status, 2005 **Features:** +: Low, Loud, Taste, Human, Complexity –: Motion, Bright, Consequential, Complexity, Human

Note: [T] Target Entity, [C] Context Entity. Entities: Role (Emotion/Property). Setting includes Place, Time, and Atmosphere. Features show top 5 positive (+) and negative (–) Binder dimensions.

Table 2: Scene Cluster Profiles and top distinguishing Binder features for IA

Illegal Aliens: Dominance of Criminal and Jurisdictional Frames The narrative landscapes for IA are predominantly anchored in themes of violation and legal institutional struggle. As shown in IA_0 (Administrative Criminalization) and IA_2 (Security Threat), the term is frequently embedded in scenes involving criminal records, border enforcement, and the legal processing of individuals. These automatically derived themes align with those of security and illegality discussed in the framing literature on immigration (Lakoff and Ferguson, 2006). Though some claim a neutral or bureaucratic meaning of the phrase IA, cluster IA_3 (Media Outrage) further indicates that IA has *resonances* (Beaver and Stanley, 2023) with, that is it conventionally co-occurs with, attitudes of *hostility* in the *speaker*.

Cluster	Scene Profiles
UI_0	**Theme: DACA & Family** **Events:** program, deportation, children, number, work, daca **Entities:** [**T**] subjects (vulnerability, uncertainty); [**C**] maker (policy, determined, hope) **Setting:** america, mexico; 2012, current; charged, tense, hopeful **Surface N-grams:** estimated 11 million, live, pew, allows, million living, lived, came children, countries, travel ban **Features:** +*:* Happy, Slow, Smell, High, Fast –*:* Disgusted, Pleasant, Large, Landmark, Pain
UI_1	**Theme: Enforcement Tension** **Events:** deportation, enforcement, authorities, agents, local **Entities:** [**T**] subjects, target (vulnerable, risk, fear); [**C**] authority (actions, policy, determined) **Setting:** border, mexico; current, admin; tense, charged **Surface N-grams:** police, arrested, arrests, city, raids, convicted, local enforcement, orders, sessions, advocates **Features:** +*:* Dark, Drive, Large, Landmark, Color –*:* Toward, Texture, Sad, Self, Temperature
UI_2	**Theme: Electoral Debate** **Events:** expresses, makes, claims, citizenship, regarding, discusses **Entities:** [**T**] subject, policy (vulnerable, legal, fear); [**C**] speaker (controversial, stance, defiant) **Setting:** campaign, election; charged, contentious, tense **Surface N-grams:** told, clinton, election, vote, cruz, position, debate, gop, presidential, speech **Features:** +*:* Unpleasant, Sad, Toward, Weight, Shape –*:* Dark, Drive, Away, Biomotion, Arousal
UI_3	**Theme: Legislative Reform** **Events:** legislation, citizenship, legal, licenses, policy, status **Entities:** [**T**] subjects, beneficiaries (status, legal, potential, hope); [**C**] decision, maker (influential) **Setting:** legislative context, california; current; charged, hopeful **Surface N-grams:** pathway, tuition, pathway citizenship, license, real id, bipartisan, medicaid, compliant **Features:** +*:* Texture, Toward, Pleasant, Surprised, Touch –*:* Dark, LowerLimb, Drive, Needs, Biomotion
UI_4	**Theme: Grassroots Advocacy** **Events:** expresses, advocates, individuals, status, identified, legal **Entities:** [**T**] subject, recipients (status, vulnerable, fear); [**C**] advocate (issues) **Setting:** unspecified, california; ongoing, present; tense, supportive **Surface N-grams:** author, contact, members, coalition, students, community, live, advocates **Features:** +*:* Path, Pattern, Pleasant, LowerLimb, Attention –*:* Happy, Harm, Cognition, Fast, Head

Note: [**T**] Target Entity, [**C**] Context Entity. Entities include Role, Property (Prop), and Emotion (Emo). Setting includes Place, Time, and Atmosphere. Features show top 5 positive (+) and negative (–) Binder dimensions.

Table 3: Scene Cluster Profiles and top distinguishing Binder features for UI

Undocumented Immigrants: Youth and Reform Frames In contrast, UI occupies a much broader and more situational spectrum. While UI_1 (Enforcement Tension) discusses similar topics to IA_0, this narrative is critical of heightened activity, painting law enforcement officers as violent. Other clusters such as UI_0 (DACA / Family) and UI_3 (Legislative Reform) shift the focus toward social protection and systemic change. Cluster UI_3 aligns with Lakoff and Ferguson (2006)'s *Reform* frame. Notably, UI_4 (Grassroots Advocacy) reveals an immigrant perspective atmosphere that is largely absent in the IA clusters. The prevalence of terms like *dreamers*, *pathway to citizenship*, and *tuition* indicates that UI is primarily used to evoke narratives of inclusion, long-term residency, and human rights.

4.3.2 Enhancing Narrative Landscapes with Cognitive Features

The LLM-derived scene descriptions and the embedding-projection derived Binder features offer complementary views of the same underlying data. Combining the tracks shows alignment between these independent methods while revealing distinctions that the scene labels alone do not capture.

We first performed one-vs-all logistic regression for each scene cluster, predicting cluster membership from Binder feature values. Tables 2 and 3 report the top five positive and negative features for each cluster ranked by regression coefficient. The features largely confirm the cluster characterizations. For IA, IA_0 (Criminalization) and IA_3 (Media Outrage) show the highest **Fearful** values, and elevated **Harm**, consistent with rhetoric depicting migrants as dangerous.

IA_3 (Security Threat) suppresses the features **Social** and **Human** and emphasizes **Disgusted**, potentially indicating dehumanization narratives. IA_4 (Jurisdictional Conflict), a more abstract narrative, emphasizes **Complexity**. Clusters UI_2 and UI_4, which discuss government and politicians decisions *about* migrants, both suppress **Biomotion**. The former, which discusses controversial election cycles, is more **Unpleasant** and **Sad**, while the latter, which discusses presents progressive reform in a positive light, is more **Pleasant**, as is UI_4, containing migrant perspectives.

Integration of both analytic methods reveals that the feature activations for each phrase are not monolithic. Though variation is driven by lexical choice, contextual variation plays role. A Kruskal–Wallis test across clusters within each phrase showed significant ($p < 0.05$) variation among 27/65 features for IA and among 59/65 features for UI (Benjamini-Hochberg corrected p-values). Figure 3 shows the mean values of the key features identified in Section 4.1 across scene clusters. For example,IA_0 Administrative Criminalization is much higher than other IA clusters on **Human**. This suggests that even within IA usage, some narrative contexts depictions are more humanizing, if negatively so. After all, getting arrested and committing crimes are uniquely human activities. When discussed as a plank in a political platform (IA_1), the phrase IA behaves more like non-human words.

As noted above, **Angry** is particularly high for UI_1 (Enforcement Tension) and IA_0 (Administrative Criminalization). They both discuss law

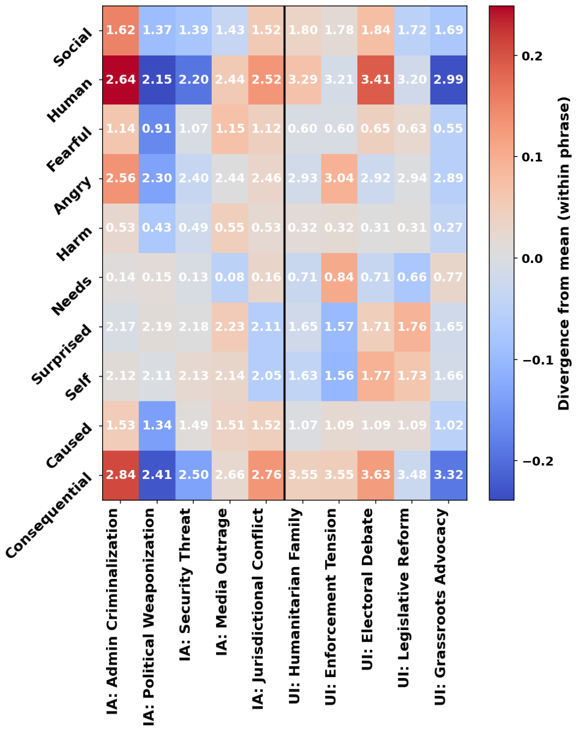

Figure 3: Mean Binder features for aliens and immigrants scene clusters, with colors showing each cluster's deviation from the mean for that phrase. Compare values across the board; compare colors within IA and UI separately.

enforcement operations, but only in UI_1 is there also a spike in **Needs**, perhaps reflecting the more migrant-oriented perspective of that cluster.

5 Discussion

Our two methods independently converged on the same finding: IA contexts emphasize immigrants as agents (via **Caused** and criminal/threat narratives), while UI contexts emphasize immigrants as experiencers of consequences (via *Consequential* and policy/enforcement narratives). This convergence validates both methods and aligns with discourse analyses of criminal and victim frames (Lakoff and Ferguson, 2006; Khosravinik, 2010; Zawadzka-Paluektau, 2023).

The criminal/victim distinction is often treated as a matter of valence. But our analysis suggests a deeper structural asymmetry involving *agency*. To probe this, we first labeled target entity roles as agentic or passive, then calculated the dominant association of each Scene Abstraction target role (IA or UI). Seven of the top ten IA roles were agentic (*criminals, laborers, voters, perpetrators, workers, agents, migrants*) versus four of the top ten UI roles (*applicants, contributors, taxpayers, participants*—themselves more abstract, less active roles). UI was dominated by passive roles like *population*, *vulnerable population*, *affected group*, and *subjects of policy*.

These results complicate a simple euphemism/dysphemism account of *undocumented immigrants* and *illegal aliens*. UI may construct frames or evoke conceptual representations that are more sympathetic, but it may also reinforce a paternalistic construal of migrants as passive recipients rather than agents navigating difficult circumstances.

6 Conclusion

We have shown that the phrases *undocumented immigrants* and *illegal aliens*, which have similar referents but are associated with different political positions, activate systematically different semantic features in naturalistic text. Partisan word choice participates in systematically different construals: patients versus agents, victims versus perpetrators. The sympathetic frame is not simply the inverse of the hostile one; it carries its own implications about who acts and who is acted upon. Beyond the substantive findings, this work demonstrates the utility of psycholinguistic feature projection and scene abstraction for computational analysis of social meaning. Together, these methods offer a low-annotation pathway from computational detection of political language to description.

Future work will further explore the relationship between narrative frames used to discuss immigration and the implications for agency, both linguistic and social. One potential approach would be to use semantic role labeling. We are especially excited about the possibilities of SAEs (Cunningham et al., 2023; Templeton et al., 2024) for semantic analysis, though they, too, face issues with interpretability. Future work will explore these methods on a large scraped dataset of immigration news. There would be value in applying the methods developed here to other domains in which narrative frames have been closely examined such as health (Bodd et al., 2022) or economics (Cai, 2021; Mohler et al., 2016), as well as extension to other, less-well understood domains of political and cultural variation.

Limitations

The Partisan News dataset is limited to pre-2019 discourse, and therefore cannot be used to examine shifts in immigration framing around the 2020 and

2024 elections.

The methodology of interpretable embeddings for social-scientific analysis is promising, but existing tools are limited. Infrastructure for creating embeddings based on Binder feature set was used in part because it is readily available (Ranganathan et al., 2025); future work would benefit from the annotation of feature-based representations of words in context, and explore the impact of different feature sets on results.

References

Marianna Apidianaki and Aina Garí Soler. 2021. ALL Dolphins Are Intelligent and SOME Are Friendly: Probing BERT for Nouns' Semantic Properties and their Prototypicality. *arXiv preprint*. ArXiv:2110.06376 [cs].

Collin F. Baker. 2014. FrameNet: A Knowledge Base for Natural Language Processing. In *Proceedings of Frame Semantics in NLP: A Workshop in Honor of Chuck Fillmore (1929-2014)*, pages 1–5, Baltimore, MD, USA. Association for Computational Linguistics.

David Beaver and Jason Stanley. 2023. *The Politics of Language*. Princeton University Press.

Jeffrey R. Binder, Lisa L. Conant, Colin J. Humphries, Leonardo Fernandino, Stephen B. Simons, Mario Aguilar, and Rutvik H. Desai. 2016. Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, 33(3-4):130–174.

Monica H. Bodd, Nikki C. Daniels, Hermioni L. Amonoo, Tyler P. Tate, Kris W. Herring, and Thomas William LeBlanc. 2022. "More than conquerors": A qualitative analysis of war metaphors for patients with cancer. *Journal of Clinical Oncology*, 40(e24134-e24134).

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In *Advances in neural information processing systems*, volume 29.

Erin M. Buchanan, K. D. Valentine, and Nicholas P. Maxwell. 2019. English semantic feature production norms: An extended database of 4436 concepts. *Behavior Research Methods*, 51(4):1849–1863.

Dongman Cai. 2021. *Corpus studies on the use and framing of metaphor in popular economic discourse*. phd, University of Leeds.

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky. 2022. Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration. *Proceedings of the National Academy of Sciences*, 119(31):e2120510119.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised learning of narrative event chains. In *Proceedings of ACL-08: HLT*, pages 789–797.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 602–610.

Emmanuele Chersoni, Enrico Santus, Chu-Ren Huang, and Alessandro Lenci. 2021. Decoding Word Embeddings with Brain-Based Semantic Features. *Computational Linguistics*, 47(3):663–698.

Yejin Cho and Katrin Erk. 2026. Scene abstraction for lexical semantics: Structured representations of situated meaning. *arXiv preprint*.

Gabiriella Chronis. 2026. *NLP as Language Ideology: Discourse and Distribution [Unpublished doctoral dissertation]*. Ph.D. thesis, The University of Texas at Austin.

Gabriella Chronis, Kyle Mahowald, and Katrin Erk. 2023. A Method for Studying Semantic Construal in Grammatical Constructions with Interpretable Contextual Embedding Spaces. In *Proceedings of the 61st Annual Meeting of the ACL(Volume 1: Long Papers)*, pages 242–261, Toronto, Canada. Association for Computational Linguistics.

J. David Cisneros. 2008. Contaminated Communities: The Metaphor of "Immigrant as Pollutant" in Media Representations of Immigration. *Rhetoric & Public Affairs*, 11(4):569–601.

Hoagy Cunningham, Aidan Ewart, Logan Riggs, Robert Huben, and Lee Sharkey. 2023. Sparse Autoencoders Find Highly Interpretable Features in Language Models. *arXiv preprint*. ArXiv:2309.08600 [cs].

Stefan Dahlberg, Sofia Axelsson, and Sören Holmberg. 2020. Democracy in context: using a distributional semantic model to study differences in the usage of democracy across languages and countries. *Zeitschrift für Vergleichende Politikwissenschaft*, 14(4):425–459.

Katrin Erk and Gabriella Chronis. 2022. Word embeddings are word story embeddings (and that's fine). In *Algebraic Structures in Natural Language*, pages 189–218. CRC Press.

Charles J. Fillmore. 1976. Frame semantics and the nature of language. In Stevan R. Harnad, Horst D. Steklis, and Jane Lancaster, editors, *Origins and evolution of language and speech*, Annals of the New York Academy of Sciences, pages 20–32. New York Academy of Sciences, New York.

Costas Gabrielatos and Paul Baker. 2008. Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics*, 36(1):5–38.

Gloria Gennaro and Elliott Ash. 2022. Emotion and Reason in Political Language. *The Economic Journal*, 132(643):1037–1059.

Matthew Gentzgow and Jesse M. Shapiro. 2010. What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.

Yuling Gu, Bhavana Dalvi, and Peter Clark. 2022. DREAM: Improving Situational QA by First Elaborating the Situation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1115–1127, Seattle, United States. Association for Computational Linguistics.

Akram Sadat Hosseini and Steffen Staab. 2024. Disambiguating Emotional Connotations of Words Using Contextualized Word Representations. In *Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM 2024)*, pages 264–277, Mexico City, Mexico. Association for Computational Linguistics.

Jena D. Hwang, Chandra Bhagavatula, Ronan Le Bras, Jeff Da, Keisuke Sakaguchi, Antoine Bosselut, and Yejin Choi. 2021. COMET-ATOMIC 2020: On Symbolic and Neural Commonsense Knowledge Graphs. *arXiv preprint*. ArXiv:2010.05953 [cs].

Majid Khosravinik. 2010. The representation of refugees, asylum seekers and immigrants in British newspapers: A critical discourse analysis. *Journal of Language and Politics*, 9(1):1–28.

Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 Task 4: Hyperpartisan News Detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Austin C. Kozlowski, Matt Taddy, and James A. Evans. 2019. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *American Sociological Review*, 84(5):905–949.

George Lakoff and Sam Ferguson. 2006. The Framing of Immigration.

Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris Mcnorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4):547–559.

Michael Mohler, Mary Brunson, Bryan Rink, and Marc Tomlinson. 2016. Introducing the LCC Metaphor Datasets. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4221–4227, Portorož, Slovenia. European Language Resources Association (ELRA).

Maria Leonor Pacheco, Tunazzina Islam, Lyle Ungar, Ming Yin, and Dan Goldwasser. 2023. Interactive Concept Learning for Uncovering Latent Themes in Large Text Collections. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5059–5080, Toronto, Canada. Association for Computational Linguistics.

James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Commissioned report, University of Texas at Austin.

Andrew Peterson and Arthur Spirling. 2018. Classification Accuracy as a Substantive Quantity of Interest: Measuring Polarization inWestminster Systems. *Political Analysis*, 26(1):120–128.

Mattia Proietti, Gianluca Lebani, and Alessandro Lenci. 2022. Does BERT Recognize an Agent? Modeling Dowty's Proto-Roles with Contextual Embeddings. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4101–4112, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jwalanthi Ranganathan, Rohan Jha, Kanishka Misra, and Kyle Mahowald. 2025. semantic-features: A User-Friendly Tool for Studying Contextual Word Embeddings in Interpretable Semantic Spaces. *Society for Computation in Linguistics*, 8(1).

Ludovic Rheault and Christopher Cochrane. 2020. Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora. *Political Analysis*, 28(1):112–133.

Maarten Sap, Ronan Le Bras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A Smith, and Yejin Choi. 2019. Atomic: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 3027–3035.

Michael Silverstein. 2009. The Limits of Awareness. In Alessandro Duranti, editor, *Linguistic Anthropology: A Reader*, 2nd edition, volume 84, pages 382–4001. Southwest Educational Development Laboratory.

Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Charlotte Taylor. 2021. Metaphors of migration over time. *Discourse & Society*, 32(4):463–481.

Adly Templeton, Tom Conerly, Jonathon Marcus, Jack Lindsey, Trenton Bricken, and Brian Chen. 2024. Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet.

Jacob Turton, Robert Elliott Smith, and David Vinson. 2021. Deriving Contextualised Semantic Features from BERT (and Other Transformer Model) Embeddings. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*, pages 248–262, Online. Association for Computational Linguistics.

Albert Webson, Zhizhong Chen, Carsten Eickhoff, and Ellie Pavlick. 2020. Are "Undocumented Workers" the Same as "illegal aliens"? Disentangling denotation and connotation in vector spaces. In *Proceedings of EMNLP*.

Natalia Zawadzka-Paluektau. 2023. Metaphors of Crisis Migration: Evidence from a Cross-National Corpus Study. *CADAAD Journal*, 15(2):124–146.

A Binder Feature Definitions and Interpretation

The Binder features used in this study derive from human ratings of 535 English words on 65 semantic dimensions (Binder et al., 2016). Participants rated each word on a 1–7 scale for each feature. To aid interpretation of the regression results, we provide the original Binder definitions, the highest-rated anchor words, and example sentences from our corpus with extreme feature values.

A.1 Interpreting Anchor Words

The anchor words reflect which concrete concepts scored highest on each dimension in the original Binder norming study. However, these anchors do not always straightforwardly predict how features behave when applied to contextualized embeddings of social categories. For instance, the **Needs** feature captures words representing things that organisms require for survival (e.g., *sun, water, mouth, hand, cash*), but high Needs values in our corpus do not necessarily indicate that immigrants are described as *having* needs—rather, the feature may activate in contexts discussing basic human requirements, vulnerability, or dependency more broadly. Similarly, features like **Sound** and **Music** have anchor words that are literal sound-producing objects (*piano, drum, banjo*), yet their predictive power in immigration discourse likely reflects something other than acoustic properties.

We note below which features yield interpretable patterns in our qualitative analysis and which remain opaque despite strong predictive power.

A.2 Features Predictive of "Undocumented Immigrants"

Consequential *Definition:* Likely to have consequences (cause other things to happen). *Anchor words:* perjury, gunshot, election, tornado, vice, cyclone, sin, hurricane, stole, grievance. High-Consequential sentences portray immigrants as experiencing significant consequences rather than causing them. Annotators characterized these contexts as depicting immigrants as subjects of consequential actions by others.

Needs *Definition:* Someone or something that would be hard for you to live without. *Anchor words:* sun, drank, water, mouth, eye, leg, lip, hand, cash, ate. Despite strong predictive power, annotators found no coherent pattern distinguishing high- and low-Needs sentences.

Human *Definition:* Having human or human-like intentions, plans, or goals. *Anchor words:* activist, diplomat, driver, pilot, journalist, soldier, woman, commander, businessman, lawyer. High-Human sentences use what annotators called "human verbs"—immigrants *know*, *seek*, *enjoy*, are *protected*, and possess rights. Low-Human sentences treat immigrants more as objects or resources; as one annotator noted, "they could be boxes."

LowerLimb *Definition:* Associated with actions using the leg or foot. *Anchor words:* hiked, kicked, walked, foot, soccer, marched, leg, ran, bicycle, toe. High-LowerLimb sentences frequently describe deportation procedures, law enforcement operations, and physical displacement—contexts involving bodily movement and institutional action upon immigrants. Low-LowerLimb sentences more often discuss funding, tuition, and structural policies.

Complexity *Definition:* Visually complex. *Anchor words:* jungle, carnival, garden, airport, zoo, circus, festival, home, lab, rocket. High-Complexity sentences tend to describe socially vulnerable people in nuanced situations. Low-Complexity sentences more often frame immigrants as recipients of benefits in simpler transactional terms.

UpperLimb *Definition:* Associated with actions using the arm, hand, or fingers. *Anchor words:* hand, handshake, keyboard, pen, finger, drew, applause, threw, wrote, carried. Like LowerLimb, high-UpperLimb sentences focus on deportation and law enforcement operations. Low-UpperLimb sentences emphasize money and economic figures ($50 million, $6 billion, $400 billion)—resources framed as being diverted to immigrants.

Speech *Definition:* Someone or something that talks. *Anchor words:* commander, diplomat, boy, businessman, teacher, actor, lawyer, mayor, minister, politician. High-Speech sentences describe immigrants in fearful or threatened states, often being detained or deported. Low-Speech sentences more often describe how immigrants are treated or provided for—as beneficiaries rather than agents.

Short *Definition:* An event that lasts for a short period of time. *Anchor words:* gunshot, handshake, gasp, belch, lightning, ricochet, dropped, thunder, screech, opened. High-Short sentences emphasize protection, rights, and dignity. Low-Short sentences focus on economic concerns—social welfare, taxpayer money, and benefits framed as going to immigrants.

A.3 Features Predictive of "Illegal Aliens"

Sound *Definition:* Having a characteristic or recognizable sound or sounds. *Anchor words:* piano, drum, bagpipe, bell, harmonica, trombone, flute, cough, accordion, gunshot. Despite being the strongest predictor of "illegal aliens," annotators found no coherent pattern. High-Sound sentences did not share obvious thematic or linguistic features.

Surprised *Definition:* Someone or something that makes you feel surprised. *Anchor words:* gunshot, explosion, scream, accident, landslide, screech, wonder, snake, clang, terrorist. High-Surprised sentences often involve congressional reform and legislation. Low-Surprised sentences describe provisions that would protect immigrants.

Self *Definition:* Related to your own view of yourself, a part of YOUR self-image. *Anchor words:* lived, worked, ate, slept, intellect, worth, drank, saw, belief, read. High-Self sentences describe immigrants voting, registering, or having children—active civic participation. Low-Self sentences describe immigrants as beneficiaries: reluctant to seek help, not given attorneys, eligible for programs.

Caused *Definition:* Caused by some clear preceding event, action, or situation. *Anchor words:* riot, honeymoon, destroyed, grievance, apology, damaged, trial, funeral, truce, battle. This feature yielded the clearest qualitative pattern. High-Caused sentences portray immigrants as agents of harm: causing wildfires, committing sexual assaults, posing threats. Low-Caused sentences reverse this entirely—immigrants are acted upon, their fates decided by others.

Fearful *Definition:* Someone or something that makes you feel afraid. *Anchor words:* feared, terrorist, tornado, dangerous, cyclone, dread, torment, gunshot, alligator, hurricane. High-Fearful sentences describe immigrants as threats to public safety, committing sexual assaults, or presenting danger. Low-Fearful sentences discuss pathways to citizenship, registration papers, and government identification.

Music *Definition:* Making a musical sound. *Anchor words:* banjo, flute, harp, piano, symphony, trombone, trumpet, mandolin, xylophone, bagpipe. Despite strong predictive power, the connection to immigration discourse is opaque. Annotators noted that high-Music sentences often involved danger, threat, and crime, but the link to the feature's original definition remains unclear.

Head *Definition:* Associated with actions using the face, mouth, or tongue. *Anchor words:* jaw, drank, gum, kiss, mouth, shouted, tuba, ate, harmonica, flute. No clear pattern emerged from qualitative analysis.

Body *Definition:* Having human or human-like body parts *Anchor words:* criminal, parent, pilot, politician, scientist, soldier, witness, worker, activist, businessman.

A.4 Example Sentences with Extreme Feature Values

Tables 4 and 5 present example sentences from our corpus that scored highest and lowest on each feature. Target phrases are shown in italics.

Table 4: Example sentences for features predictive of "undocumented immigrants" (positive β). High-scoring sentences appear in the left column; low-scoring sentences appear in the right column.

Feature	High-Scoring Sentences	Low-Scoring Sentences
Consequential	A memo released Tuesday by Homeland Security provided further details on how the Trump administration will carry out immigration enforcement, with the new guidance essentially allowing the deportation of many more *undocumented immigrants* through expedited removal. Today we stand together with the 800,000 young people who will lose their status as a result of DACA's termination as well as the 11 million *undocumented immigrants* who have watched this president unleash his deportation force upon them with impunity. While *undocumented immigrants* in particular take on many of the most dangerous jobs in society, their wages are often far below that other workers. It's not a benign observation, but has meant tragic consequences for many people in our country—from terror-inducing raids in the communities of *undocumented immigrants* to his disparaging of refugees.	In addition, *illegal aliens* may "not be permitted to enroll in or attend any public postsecondary education institution." New Mexico has at least three detention facilities specifically designated to hold *undocumented immigrants*, one each in Torrance, Cibola and Otero counties. Through a deliberate lack of prosecution, President Obama has made America a sanctuary nation for felons, criminal gangbangers, drug dealers, repeat offenders and *illegal aliens*. Bannon, a Catholic, said the church hasn't come to grips with its own larger problems and as a result "they need *illegal aliens* to fill the churches."
Human	While *undocumented immigrants* in particular take on many of the most dangerous jobs in society, their wages are often far below that other workers. They are slowing deportations by insisting that *undocumented immigrants* still have the right of due process, even if in many of these cases, the immigrants had known for years that they could be expelled. The other is that 11 million *undocumented immigrants* continue to enjoy a de facto amnesty while our society remains less productive and less secure than it should be. One of the most consistently newsworthy developments have been the Dream Activists: young *undocumented immigrants* seeking to enforce the United Nations-declared universal human right to a nationality.	In addition, *illegal aliens* may "not be permitted to enroll in or attend any public postsecondary education institution." Most recently, the ACLJ successfully defended the right of states to impose business license restrictions on companies that knowingly hire *illegal aliens*. Well, it didn't take Wilson or any other conservative critic a crystal ball to see that Democrats would be eager to dole out American taxpayer money to *illegal aliens*. Bannon, a Catholic, said the church hasn't come to grips with its own larger problems and as a result "they need *illegal aliens* to fill the churches."
Short	The program, established by Barack Obama in 2012, protects *undocumented immigrants* who entered the country as children, from being deported. For the last five years, Ronquillo, like 800,000 other previously *undocumented immigrants*, has been protected from deportation thanks to DACA. While *undocumented immigrants* in particular take on many of the most dangerous jobs in society, their wages are often far below that other workers. The idea that *undocumented immigrants* are criminals is a pernicious one; it portrays US citizens as dignified human beings with full political and economic rights.	Bannon, a Catholic, said the church hasn't come to grips with its own larger problems and as a result "they need *illegal aliens* to fill the churches." Could we get a poll on that: Should the government issue work permits to *illegal aliens* and give them each $25,000 in U.S. taxpayer money? Well, it didn't take Wilson or any other conservative critic a crystal ball to see that Democrats would be eager to dole out American taxpayer money to *illegal aliens*. Daryl Metcalfe's State Legislators for Legal Immigration works to eliminate "all economic attractions and incentives for *illegal aliens*, as well as securing our borders against unlawful invasion."
UpperLimb	In a raid on six Swift meatpacking plants Immigration and Customs Enforcement (ICE) agents detained 1,300 Latino *undocumented immigrants* workers, charging some with identity theft and deporting others. On Sunday, the Trump administration demanded that Congress overhaul the US asylum system as part of any legislation to protect the nearly 700,000 *undocumented immigrants* known as Dreamers from deportation. Earlier this week, the Justice Department repeated a threat to withhold federal grant money to cities that actively try to subvert efforts to deport *undocumented immigrants*. Democrats also are demanding protections from deportation for *undocumented immigrants* who were brought here illegally as children.	Mexico is investing $50 million to help *illegal aliens* break US law. Daryl Metcalfe's State Legislators for Legal Immigration works to eliminate "all economic attractions and incentives for *illegal aliens*." For many years they were intentionally breaking the law by hiring thousands of *illegal aliens* instead of Americans. Billions of dollars will be spent resettling *illegal aliens*.
LowerLimb	The effect of his dragnet on *undocumented immigrants* is not immediately apparent. "Expedited removal" is the term the government uses to describe the swift deportation of *undocumented immigrants* without an appearance before an immigration judge. Clinton said Sanders "voted in the House with hard-line Republicans for indefinite detention for *undocumented immigrants*, and then he sided with those Republicans to stand with vigilantes known as Minutemen." Trump said that another order, greatly expanding the number of *undocumented immigrants* prioritized for deportation, meant the start of a "military operation."	But when David Duke starts speaking: "Either you get the INS to kick the *illegal aliens* out, or you'll lose your community and your heritage." Pre-Senate Run: As a Florida state legislator, Rubio supported legislation that would have provided in-state tuition to *illegal aliens*. What would you say if the Obama administration awarded a $50 million contract to buy a Texas resort hotel and transform it in to a 600 bed facility for juvenile *illegal aliens*. "$2.2 Billion dollars a year is spent on food assistance programs such as food stamps, WIC, and free school lunches for *illegal aliens*."

Table 5: Example sentences for features predictive of "illegal aliens" (negative β). High-scoring sentences appear in the left column; low-scoring sentences appear in the right column.

Feature	High-Scoring Sentences	Low-Scoring Sentences
Caused	Among this cohort of *illegal aliens* were violent criminals responsible for 193 homicide convictions, 426 sexual assault convictions, 1075 aggravated assault convictions, 16,070 drunk driving convictions.	There are 11 million *undocumented immigrants* in the country and while many are now watching from afar as their fates are decided in the nation's capital.
	A recent independent audit found all the charges against the programme to be false. "*Illegal aliens*" causing Arizona wildfires?	The piece has been installed during a particularly fraught time for U.S.-Mexico relations, as the Trump administration repeatedly threatens *undocumented immigrants*.
	In 2011 the Government Accountability Office reported that *illegal aliens* incarcerated in US prisons in 2010 were responsible for seventy thousand sexual assaults.	It also reinforces the workings of a corporate-driven culture whose airwaves are filled with hate, endless spectacles of violence and an ongoing media assault on young people, the poor, Muslims and *undocumented immigrants*.
	During the struggle the two *illegal aliens* attempted to choke the agent.	As organizers, we must focus less on meaningless negotiations, and more on building power and leadership among those impacted—namely *undocumented immigrants*.
	We will continue collaborating with them to help ensure that *illegal aliens* who may pose a threat to our communities are not released onto the streets.	A lesbian activist posed a question to Donald Trump Wednesday night on *undocumented immigrants*.
Fearful	In 2011 the Government Accountability Office reported that *illegal aliens* incarcerated in US prisons in 2010 were responsible for seventy thousand sexual assaults.	Some are *undocumented immigrants*, brought here as children through no fault of their own.
	Lest the *illegal aliens* become nervous, however, the legislation tells them they need not be.	Cornyn's amendment to the immigration reform bill would essentially have tied the building of a border fence to the ability of *undocumented immigrants* to obtain green cards.
	The film makes clear that the aliens are the "worker bees" whose sole function is to perform labor, not unlike the masses of *illegal aliens* who work in the sweat factories.	Like the Senate group's immigration reform package, the president's proposal would provide a pathway to citizenship for the nation's estimated 11 million *undocumented immigrants*.
	These policies also encourage illegal immigration and even human trafficking by perpetuating the lie that in certain cities, *illegal aliens* can live outside the law.	When Trump talks about *undocumented immigrants*, he's talking about Latinos, and he targets in particular Mexicans, right?
	After his inauguration, the discourse took on a new form as a dehumanizing bureaucratic language, casting people without papers as "*illegal aliens*" who "present a significant threat to national security."	A lesbian activist who pushed for "Don't Ask, Don't Tell" repeal posed a question to Donald Trump Wednesday night on *undocumented immigrants*.
Sound	What would you say if the Obama administration awarded a $50 million contract to a charitable group to buy a Texas resort hotel and transform it in to a 600 bed facility for juvenile *illegal aliens*.	The British Home Office said the majority of the detainees are failed asylum seekers or *undocumented immigrants* waiting to be transported to their homelands.
	But when David Duke starts speaking: "Either you get the INS to kick the *illegal aliens* out, or you'll lose your community and your heritage."	As organizers, we must focus less on meaningless negotiations, and more on building power and leadership among those impacted—namely *undocumented immigrants*.
	The idea was to prevent *illegal aliens* from claiming any federal or state benefit that wasn't first made available to legal residents.	Why is it acceptable for *undocumented immigrants* to live in constant fear, for their families to be at constant risk of being torn apart?
	Supporters of comprehensive immigration reform scored a major victory when the California Supreme Court ruled that *illegal aliens* could qualify for in-state tuition.	Rep. Michele Bachmann took to Fox News and called on tea partiers everywhere to come to Washington to protest Obama's executive action that would protect millions of *undocumented immigrants*.
	And Dan Scavino, Jr. tweeted: "Terrible. We know who the 1,000+ *illegal aliens* ARE NOT VOTING FOR!"	One of the Legislature's first acts was an attempt to make good on Martinez's pledge to revoke driver's licenses for *undocumented immigrants*.
Self	Pre-Senate Run: As a Florida state legislator, Rubio supported legislation that would have provided in-state tuition to *illegal aliens*.	If Mexico refuses, then the United States will impound remittance payments taken from the wages of *undocumented immigrants*.
	But when David Duke starts speaking: "Either you get the INS to kick the *illegal aliens* out, or you'll lose your community and your heritage."	Officials emphasized that identification and legal status are not required at the shelters, after reports that *undocumented immigrants* were reluctant to seek help.
	Three million votes in the U.S. presidential election were cast by *illegal aliens*, according to Greg Phillips.	Clinton said Sanders "voted in the House with hard-line Republicans for indefinite detention for *undocumented immigrants*."
	This in itself negates the votes made by *illegal aliens* in California and any other state which allows them to vote.	The policy permits *undocumented immigrants* who were brought to the United States by their parents to apply for a two-year residency.
	Voter ID laws are not about preventing *illegal aliens* or other unauthorized people from registering to vote.	Santa Clara County supervisors agreed to invest about $1.5 million toward legal aid for *undocumented immigrants* in danger of deportation.

B Scene Abstraction

B.1 Top Predictor Features of UI/IA Labels

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[T] role: population`	3.06
`[T] role: vulnerable population`	2.35
`[T] role: subject of discourse`	1.76
`[T] role: applicants`	1.49
`[T] role: affected group`	1.35
`[T] role: contributors`	1.14
`[T] role: beneficiaries`	0.99
`[T] role: target`	0.91
`[T] role: subject of claim`	0.87
`[T] role: affected population`	0.83
Top Predictors for Illegal Aliens (IA)	
`[T] role: criminals`	-1.15
`[T] role: recipients`	-0.89
`[T] role: subjects of enforcement`	-0.82
`[T] role: laborers`	-0.80
`[T] role: voters`	-0.52
`[T] role: victims`	-0.52
`[T] role: subjects of concern`	-0.42
`[T] role: subject of policy`	-0.42
`[T] role: perpetrators`	-0.39
`[T] role: workers`	-0.37

Note: Prefix is abbreviated as: **[T]**: Target Entity.

Table 6: Top 10 Predictive **Target Entity Role** Features for UI and IA Labels.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[T] prop: young`	2.38
`[T] prop: vulnerable status`	2.18
`[T] prop: seeking legal recognition`	2.16
`[T] prop: lacking legal status`	2.05
`[T] prop: resilient`	1.49
`[T] prop: large demographic`	1.46
`[T] prop: vulnerable`	1.45
`[T] prop: long-term residents`	1.42
`[T] prop: seeking legal status`	1.29
`[T] prop: law-abiding`	1.15
Top Predictors for Illegal Aliens (IA)	
`[T] prop: undocumented`	-4.28
`[T] prop: assoc w/ criminality`	-2.36
`[T] prop: unauthorized`	-2.19
`[T] prop: undocumented status`	-2.00
`[T] prop: unlawful status`	-1.81
`[T] prop: mobile`	-1.69
`[T] prop: controversial`	-1.46
`[T] prop: classified as illegal`	-1.38
`[T] prop: stigmatized`	-1.24
`[T] prop: non-citizens`	-0.84

Note: Prefix is abbreviated as: **[T]**: Target Entity.

Table 7: Top 10 Predictive **Target Entity Property** Features for UI and IA Labels.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[C] role: critics`	1.58
`[C] role: respondents`	1.33
`[C] role: stakeholders`	1.22
`[C] role: host`	0.96
`[C] role: political figure`	0.89
`[C] role: policy`	0.87
`[C] role: providers`	0.80
`[C] role: political leader`	0.76
`[C] role: researcher`	0.74
`[C] role: collective`	0.72
Top Predictors for Illegal Aliens (IA)	
`[C] role: destination`	-1.27
`[C] role: perpetrators`	-1.11
`[C] role: government agency`	-0.99
`[C] role: migrants`	-0.88
`[C] role: endorser`	-0.63
`[C] role: medium`	-0.59
`[C] role: data collector`	-0.52
`[C] role: respondent`	-0.47
`[C] role: law enforcement`	-0.45
`[C] role: critic`	-0.41

Note: Prefix is abbreviated as: **[C]**: Context Entity.

Table 8: Top 10 Predictive **Context Entity Role** Features for UI and IA Labels.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[C] prop: regulatory`	1.22
`[C] prop: vulnerable population`	1.00
`[C] prop: vulnerable`	0.99
`[C] prop: supportive of undocumented immigrants`	0.89
`[C] prop: vulnerable status`	0.75
`[C] prop: compliant`	0.60
`[C] prop: non-citizen status`	0.55
`[C] prop: supportive`	0.50
`[C] prop: governmental`	0.45
`[C] prop: conservative`	0.39
Top Predictors for Illegal Aliens (IA)	
`[C] prop: criminal`	-2.03
`[C] prop: divided`	-1.45
`[C] prop: aggressive`	-1.44
`[C] prop: undocumented status`	-0.98
`[C] prop: empathetic`	-0.96
`[C] prop: organized`	-0.89
`[C] prop: concerned`	-0.77
`[C] prop: citizens`	-0.73
`[C] prop: advocacy group`	-0.67
`[C] prop: manipulative`	-0.59

Note: Prefix is abbreviated as: **[C]**: Context Entity.

Table 9: Top 10 Predictive **Context Entity Property** Features for UI and IA Labels.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[T] emotion: resilience`	2.59
`[T] emotion: hope`	2.10
`[T] emotion: anxiety`	1.09
`[T] emotion: vulnerability`	0.91
`[T] emotion: fear`	0.86
`[T] emotion: uncertainty`	0.83
`[T] emotion: marginalized`	0.78
`[T] emotion: insecurity`	0.76
`[T] emotion: anxious`	0.73
`[T] emotion: hopeful`	0.71
Top Predictors for Illegal Aliens (IA)	
`[T] emotion: disempowered`	-0.42
`[T] emotion: controversial`	-0.17
`[T] emotion: none`	-0.17
`[T] emotion: uncertain`	-0.10

Note: Prefix is abbreviated as: **[T]**: Target Entity.
Less than 10 Target Entity Emotion predictors found for IA.

Table 10: Top 10 Predictive **Target Entity Emotion** Features for UI and IA Labels.

Feature Descriptor (Abbreviated)	**Coeff.** (β)
Top Predictors for Undocumented Immigrants (UI)	
`[C] emotion: hope`	1.09
`[C] emotion: relief`	0.89
`[C] emotion: anxiety`	0.69
`[C] emotion: conviction`	0.62
`[C] emotion: vulnerability`	0.57
`[C] emotion: resolute`	0.56
`[C] emotion: anxious`	0.43
`[C] emotion: ambivalence`	0.36
`[C] emotion: hopeful`	0.36
`[C] emotion: disapproval`	0.25
Top Predictors for Illegal Aliens (IA)	
`[C] emotion: disdain`	-1.00
`[C] emotion: controversial`	-0.96
`[C] emotion: indifference`	-0.84
`[C] emotion: urgency`	-0.62
`[C] emotion: frustration`	-0.49
`[C] emotion: pragmatic`	-0.32
`[C] emotion: indignant`	-0.28
`[C] emotion: anger`	-0.25
`[C] emotion: critical`	-0.25
`[C] emotion: defensive`	-0.21

Note: Prefix is abbreviated as: **[C]**: Context Entity.

Table 11: Top 10 Predictive **Context Entity Emotion** Features for UI and IA Labels.

B.2 Full Scene Abstraction Example

```
[Input Sentence]
"He urged a balanced approach, noting that those
     undocumented immigrants who commit serious crimes
     'have no business in the United States.'"

[Scene Abstraction]
{
  "Tier_1_Global_Scene": {
    "Events": [
      "PersonX urges a balanced approach",
      "PersonX notes the status of GroupY",
      "PersonX states that GroupY who commit
           serious crimes should not be in the US"
    ],
    "Entities": [
      {
        "Entity": "PersonX (He)",
        "Roles": ["Advocate (Promoting_approach)"],
        "Properties": ["Reasoned (Thoughtful
             perspective on immigration)"],
        "Emotions": ["Concern (Emphasis on balanced
             approach)"]
      },
      {
        "Entity": "GroupY (undocumented immigrants)",
        "Roles": ["Subject (Being_discussed)"],
        "Properties": [
          "Vulnerable (Lack of legal protection)",
          "Diverse (Varying backgrounds and
               circumstances)"
        ],
        "Emotions": ["Fear (Potential consequences
             of being labeled)"]
      }
    ],
    "Setting": {
      "Place": "unspecified (Specificity: low)",
      "Time": "unspecified (Specificity: low)",
      "Atmosphere": "serious and contentious
           (Tense atmosphere)"
    }
  },
  "Tier_2_Local_Scene": {
    "Keyword": "undocumented
         immigrants (Label: GroupY)",
    "Engaged_Events": [
      "PersonX discusses their status",
      "PersonX highlights their involvement in
           crime"
    ],
    "Generalizable_Properties": [
      "Subjects of political and social debate",
      "Perceived as a mixed group with both
           positive and negative associations"
    ],
    "Evoked_Emotions": [
      "Concern (Surrounding legal status and
           societal impact)",
      "Tension (Due to the serious nature of the
           crimes mentioned)"
    ]
  }
}
```

Figure 4: Example of a full structured Scene Abstraction.

B.3 Scene Cluster Visualization

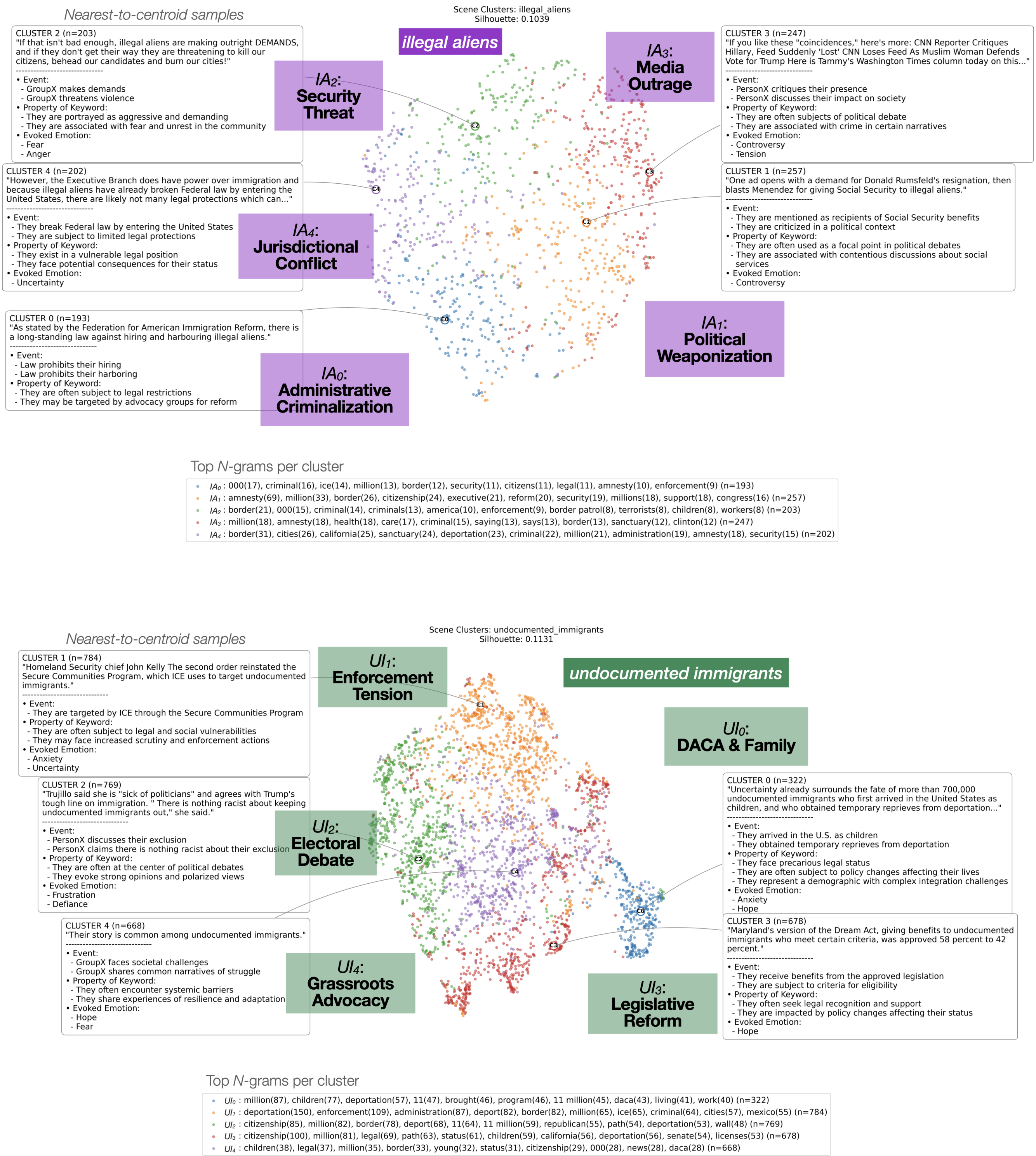

Figure 5: UMAP Visualizations of Scene Clusters.
(Top: *illegal aliens*, Bottom: *undocumented immigrants*)