Toward Unsupervised Conceptual Metaphor Discovery: A Case Study in Online Immigration Discourse

Alexandria Leto
University of Colorado Boulder
alexandria.leto@colorado.edu

Maria Leonor Pacheco
University of Colorado Boulder
maria.pacheco@colorado.edu

Abstract

In Conceptual Metaphor Theory (CMT), a metaphor is a systematic mapping from a concrete source domain (e.g., physical load) to a more abstract target domain (e.g., taxes), so that reasoning about concepts in the target domain is guided by inferences from the source domain (Lakoff, 1993). In this work, we propose that since different source domains can frame the same target in starkly different ways, the conceptual mappings evidenced by metaphorical expressions can guide computational political discourse analysis. We present a proof-of-concept for an unsupervised method that uncovers salient conceptual mappings from a corpus. Prior work in computational political metaphor analysis has drawn on CMT, but it typically requires a predetermined inventory of focused source and target domains. In contrast, we introduce a simple LLM-based method that detects metaphorical expressions from a corpus with strong performance, then clusters them to approximate source domain categories. We demonstrate its utility through a case study on online immigration discourse, showing that the resulting metaphor clusters provide context for frame analysis. We conclude by outlining future work needed to develop a robust framework for conceptual metaphor discovery in political discourse.

Code

1 Introduction

In Conceptual Metaphor Theory (CMT), a theoretical framework in cognitive linguistics that originated with Lakoff and Johnson (1980), metaphors are viewed as a fundamental structure of human thought. They allow us to understand abstract or unfamiliar concepts (the target domain) in terms of more concrete or familiar ones (the source domain). For example, in the metaphorical expression "Taxes burden the middle class," the source word "burden" invokes the *physical load* source domain (Figure 1).

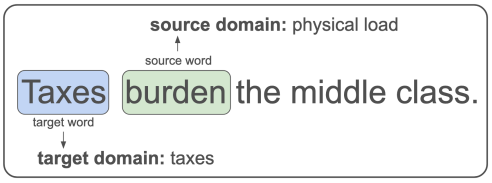

Figure 1: Example of a metaphorical expression inspired by Lakoff (2004).

Correspondences between concepts in the source domain and concepts in the target domain, referred to as conceptual metaphors or conceptual mappings (Lakoff, 1993), are thus formed: taxes are a physical load and the middle class is the carrier.

The implications of these mappings provide an understanding of how authors frame an issue, defined as "selecting some aspects of a perceived reality and making them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation" (Entman, 1993). In our example, the problem definition is the economic "burden" that the middle class is "carrying," with the implied solution being to "lift" it through reduced taxation. Conceptual metaphors therefore offer a useful lens for political framing analysis. In this work, we propose an unsupervised framework for discovering them in a corpus, recovering the metaphorical expressions used and the conceptual mappings they are drawn from.

While prior frameworks have used CMT as the scaffolding for large-scale discourse analyses, they require a pre-defined list of expected source concepts (Mendelsohn et al., 2020; Card et al., 2022; Mendelsohn and Budak, 2025). The source concepts explored are generally focused on a relatively narrow view of metaphor, honing in on "dehumanizing" metaphors, in which a person or group of people is the target concept, and the source concept (such as "vermin" or "animals") serves to strip the

Proceedings of the 7th Workshop on Natural Language Processing and Computational Social Science (NLP+CSS), pages 159–175

group of human qualities. Additionally, the target is often fixed to a particular group such as "immigrants" (Wang, 2024), precluding the discovery of other targets that may be equally relevant. For example, in immigration discourse, one might also be interested in how politicians, law enforcement, or immigration policy are metaphorically framed.

Our framework, in contrast, is more flexible than prior approaches, enabling the discovery of conceptual metaphors in novel datasets or discourse types for which no established inventory of conceptual mappings exists. This paper serves as a proof of concept for its feasibility and usefulness in political framing analysis. Below, we outline our main contributions.

1. An unsupervised method for discovering conceptual metaphors that first extracts candidate word pairs with grammatical relationships associated with metaphor use, identifies which pairs invoke a metaphor, generates rich textual descriptions for the resulting *(source word, target word)* pairs, and then clusters these descriptions so that expressions sharing a conceptual mapping are grouped together.
2. A case study demonstrating that the discovered conceptual mappings provide signal for frame prediction and reveal how immigration is framed in in online political discourse.

These contributions constitute a proof-of-concept toward an automated conceptual mapping discovery framework for frame analysis. We close with a discussion of the main challenges and opportunities for fully realizing this vision.

2 Background

Metaphor In their book *Metaphors We Live By*, Lakoff and Johnson (1980) introduced CMT, arguing that metaphor is not limited to figurative language, but is instead a fundamental structure undergirding our conceptual system. Metaphors, also referred to as conceptual metaphors or conceptual mappings, are a sets of ontological correspondences between concepts in a target domain and concepts in a source domain, allowing patterns of inference in the source domain to be applied to the target (Lakoff, 1993). The theory spurred an influx of work that, over the years, built on, criticized, and altered the original theory (Grady, 1997; Charteris-Black, 2016; Kövecses, 2017, 2000, 2020).

With this growing volume of metaphor work in linguistics, Group (2007) noted a resulting "[v]ariability in intuitions, and lack of precision about what counts as a metaphor." In response, they developed the Metaphor Identification Procedure (MIP), a step-by-step guide for identifying metaphorical lexical unit in text. In MIP, an annotator first establishes a lexical unit's meaning in the given context, then determines whether the unit has an alternative "basic" meaning (one that is more concrete, embodied, precise or historically older) in other contexts. If it does, the lexical unit is metaphorical. We adopt the operationalization of Lakoff and Johnson (1980)'s conceptual metaphors from the Language Computer Corporation (LCC) dataset (Mohler et al., 2016). In LCC, a metaphorical expression is a two-term unit within a sentence consisting of a lexical unit invoking a target domain and another invoking a source domain, and both novel and conventionalized metaphors are included. We also use LCC to benchmark our metaphor expression identification component.

Computational Metaphor Detection and Interpretation Early computational work for detecting and interpreting metaphors (Fass, 1991; Martin, 1990; Narayanan, 1999; Feldman and Narayanan, 2004; Agerri et al., 2007) relied domain-specific, hand-annotated metaphor knowledge bases such as the Master Metaphor List (Lakoff et al., 1991), MetaBank (Martin, 1994), and the Mental Metaphor Databank (Agerri et al., 2007), among others. While these hard-coded approaches were powerful, they lacked broad coverage (Shutova, 2010). As a result, practitioners turned to unsupervised methods for metaphor processing such as clustering (Mason, 2004; Shutova, 2010; Shutova and Sun, 2013; Mohler et al., 2013; Shutova et al., 2017) and topic modeling (Heintz et al., 2013), sometimes supplementing these approaches with knowledge from broader lexical resources such as WordNet (Miller, 1994).

Since, practitioners have approached metaphor analysis using supervised neural network and BERT-based models (Do Dinh and Gurevych, 2016; Swarnkar and Singh, 2018; Pedinotti et al., 2021; Li et al., 2024). However these approaches, like early work, tend to require large hand-labeled datasets for training and do not necessarily generalize well to out-of-domain data (Yang et al., 2023).

More recently, Large Language Models (LLMs) have been assessed for their ability to detect metaphors in an unsupervised setting (Dankin et al., 2022; Ichien et al., 2024; Tong et al., 2024;

Sanchez-Bayona and Agerri, 2025). For example, Puraivan et al. (2024) evaluate a set of prompts for obtaining a binary metaphorical classification for a single Spanish verb, yielding high accuracy on their test set. Tian et al. (2024) propose an explainable approach to word-level binary metaphor detection, in which they guide the LLM's reasoning with CMT-based "scaffolding." Fuoli et al. (2025) formulate the task at the phrase level, prompting LLMs to identify metaphorical phrases in movie reviews. They evaluate a range of methods, including Retrieval Augmented Generation (RAG), prompt engineering, and model fine-tuning for their ability to improve performance. This body of work shows that LLMs represent a promising avenue for unsupervised metaphor detection. We follow this line of work, using an LLM to identify metaphorical expressions and generate rich descriptions for them, then cluster these descriptions to recover conceptual mappings from a corpus.

Political Metaphor and Frame Analysis Metaphor is a powerful tool for framing political issues, as the source domain chosen emphasizes certain aspects of a target while backgrounding others (Entman, 1993; Lakoff, 2004). For example, when immigrants are described as "flooding" the border, the evoked *water* source domain casts them as an uncontrollable natural force, foregrounding threat and overwhelm. When the same target is described as "seeking refuge," a very different source domain (one of vulnerability and safety) is evoked, inviting sympathy instead.

A substantial body of computational work has operationalized this insight. Mendelsohn et al. (2020) and Card et al. (2022) each include "dehumanizing metaphor dimensions" in their respective large-scale framing analyses. Wang (2024) presents a method for extracting metaphors from politically-charged news articles by filtering word pairs which can grammatically invoke a metaphor, then using a fine-tuned RoBERTa classifier to predict a "metaphor score." Most recently, Mendelsohn and Budak (2025) released a framework for obtaining scores for seven metaphor categories associated with immigration (including "water" and "war"). However, most of these approaches rely on a predetermined set of source and target concepts (Mendelsohn et al., 2020; Card et al., 2022; Mendelsohn and Budak, 2025), restricting their analyses to metaphors that are already well-documented for a particular topic. While Wang (2024) addresses this with an unsupervised method designed to uncover new source concepts from metaphorical verbs, they limit their consideration to target nouns referencing immigrants. We present a first step toward addressing these limitations with an unsupervised, automated framework that generalizes to new topics without a pre-conceived notion of expected target or source concepts.

3 Inducing Metaphors

In CMT, metaphorical expressions in natural language serve as evidence for the conceptual mappings which underlie them. Building on this, we present an unsupervised method for identifying metaphorical expressions, paired with a clustering approach to group those that share a conceptual mappings. Following Mohler et al. (2016), we treat metaphorical expressions as within-sentence `(source word, target word)` pairs that grammatically invoke a metaphor, where the source word has an alternative "basic" meaning (one that is more concrete, embodied, precise or historically older) in other contexts (Group, 2007).

Although other parts of speech can grammatically invoke a metaphor, we follow prior work (Shutova et al., 2010) in limiting our scope to metaphors composed of a source verb and a target noun. This choice serves three purposes. First, it makes our framework more computationally tractable. Second, focusing on target nouns aligns our work with prior work on politically-charged dehumanizing metaphors (whose targets are typically nouns), giving us a meaningful point of comparison. Third, limiting source words to verbs lets us take advantage of recent work demonstrating high LLM performance on identifying metaphorical verbs (Puraivan et al., 2024).

3.1 Extracting Candidate Metaphorical Expressions

Prior work shows that metaphors occur in a limited set of grammatical patterns (Petruck and Dodge, 2016). We use this knowledge to extract a set of candidate `(source verb, target noun)` pairs from each sentence in a given corpus. We follow the procedure in Wang (2024), identifying all `(verb, noun)` pairs in each sentence, then use the *spaCy* python library[1] to determine the shortest dependency path (SDP) between each pair. If the SDP matches one of the pre-defined metaphor-

[1] `https://spacy.io/`

ical construction patterns (shown in Table 7), the corresponding target and source word pair are considered a metaphor candidate.

3.2 Metaphorical Expression Detection

To confirm whether a target and source word pair are in fact metaphorical, we prompt an LLM in a zero-shot setting to generate a "metaphor salience score" between 0 and 1, representing how likely a human would be to recognize the source verb as metaphorical in its sentence context. Because LLMs have been shown to over-classify metaphors in binary classification settings (Hicke and Kristensen-McLachlan, 2024), the continuous scores let us calibrate predictions to the task. We also prompt the LLM to generate a rich natural-language explanation for each identified metaphor, which we later use to cluster expressions with similar conceptual mappings. The prompts for this step are adapted from Puraivan et al. (2024) and can be found in the Appendix, Figure A.1.

3.3 Grouping Metaphors of Similar Domains

Source Domain Groups To identify groups of metaphors likely to invoke the same source domain, we cluster the LLM-generated explanations of each metaphorical expression (Table 1) using the K-means algorithm with cosine similarity. We embed each metaphor explanation with `SBERT` using the `all-MiniLM-L6-v2` model (Reimers and Gurevych, 2019). While this model was originally designed to produce sentence-level embeddings, it has been successfully used to produce document-level embeddings (Zhang et al., 2025). We test a variety of cluster counts, incrementing by 25 where the minimum is 25 and the maximum is 300.

Target Domain Groups To identify metaphors that map to the same target domain, we first establish a set of canonical target domains through a human-in-the-loop process. We begin by categorizing each metaphorical target noun as a "Person," "Place," or "Thing" using a zero-shot LLM prompt. Within each category, we embed target nouns using SBERT (`all-MiniLM-L6-v2`) and apply K-means to cluster semantically similar nouns, selecting k using the elbow method. We then hand-annotate the resulting groups to arrive at a canonical set of target domains per noun category, producing a hierarchical set of canonical target families such as those shown in Table 4. Finally, we prompt an LLM in a zero-shot setting to map each target noun in context to its canonical target domain. All prompts are provided in Appendix A.4.

4 Evaluation

In this section, we introduce the datasets used for evaluation and assess the quality of each pipeline component, including candidate extraction, metaphor detection, and source and target grouping.

4.1 Datasets

We evaluate our metaphor detection component (Section 3.2) using a portion of the English LCC dataset (Mohler et al., 2016). The LCC is a collection of general-domain excerpts annotated with source and target span tags and a metaphoricity score between 0 and 3 (inclusive). Each excerpt is also associated with target domains spanning a wide range of topics such as "guns," "migration," and "abortion". The full dataset contains $\sim 78,000$ annotated excerpts. However, because our method centers on identifying metaphorical (`source verb`, `target noun`) pairs, we focus on instances where the target and source spans include only a single word and where the source span is a verb, and the target span a noun. This results in a set of $\sim 10,000$ excerpts, with $5,733$ postive and $4,285$ negative examples. The distribution of metaphor scores for this set is shown in Figure 6. We include additional information, including a full list of target domains and the metaphoricity score distribution, in Appendix A.2.

To evaluate the quality of the resulting metaphor groups (Section 3.3), we use a dataset of English tweets about immigration released by Mendelsohn and Budak (2025). This dataset has two splits; the first split contains $\sim 1,600$ tweets annotated with tweet-level metaphoricity scores between 0 and 1 and a label mapping it to one of eight source domains (animal, commodity, parasite, pressure, vermin, war, water, and domain-agnostic). The second split contains $\sim 35,000$ tweets that do not include labels for metaphoricity.

4.2 Quality of Candidate Pairs

We recruit three graduate students to annotate 100 metaphor (`source verb`, `target noun`) pairs automatically extracted from the LCC dataset to evaluate the quality of our candidate extraction. Two are computer scientists and the third is a linguist. Annotators were instructed to determine whether

Score	Explanation
0.0	*enslaved* is used literally. It refers to the historical reality of African slaves in European Colonies.
0.7	*used* is applied metaphorically to the concept of *anti-blackness*, framing it as a tool.
1.0	*support* is applied metaphorically, framing a *political stance* as structural reinforcement.

Table 1: Examples of Qwen 3 metaphoricity scores and explanations.

	LCC Score Thresholds		
	1	**2**	**3**
Random	0.495	0.496	0.499
Llama3.2	0.590	0.603	0.619
Qwen3	0.794	0.783	0.781

Table 2: ROC-AUC scores for our metaphoricity scores with different LCC metaphoricity score thresholds used to determine positive samples.

		precision	**recall**	**f1**
Random	thresh = 0.2	0.570	0.792	0.663
	thresh = 0.3	0.568	0.690	0.623
RoBERTa	5-fold	0.817	0.809	**0.812**
	generalize	0.780	0.785	0.781
Llama3.2	binary prompt	0.580	0.991	0.732
	score prompt	0.588	**0.975**	0.733
Qwen3	binary prompt	**0.828**	0.752	0.788
	score prompt	0.802	0.782	0.792

Table 3: Results of baseline and our models on the LCC dataset. Our metaphor score prompt with `Qwen3` outperforms all other models.

each pair constitutes a valid metaphor candidate. Pairs were doubly-annotated with a third annotator breaking ties. We obtained a Krippendorf's α of 0.893 and determined an accuracy of 0.891.

4.3 Quality of Metaphor Identification

We evaluate the quality of our metaphor identification component using the LCC dataset. We assess both the correlation of our continuous metaphoricity scores with human annotations and the performance of our method on a binary classification task, comparing against supervised methods and a random baseline.

Metaphoricity Thresholds A key advantage of our continuous metaphoricity score is its flexibility. By adjusting the threshold, practitioners can tune the system toward capturing subtle metaphorical language or restricting it to only the most salient instances. We evaluate this claim by computing the Spearman correlation between our continuous metaphoricity scores and the hand-annotated four-point scores provided with the LCC dataset. We find that `Qwen3-8B` (Team, 2025) produces scores that correlate moderately well with human judgements ($\rho = 0.564$), while `Llama3.2-3B` (Dubey et al, 2024) shows only weak correlation ($\rho = 0.211$). This suggests that our continuous score is informative when paired with a capable model.

We also report ROC-AUC scores at various LCC metaphoricity score thresholds (Table 2). We find that `Qwen3` salience scores are moderately predictive of the binary metaphor label at every LCC threshold, significantly outperforming both `Llama3.2` and the random baseline. This supports our claim that practitioners can set a threshold suited to their task.

Binary Classification Since our scores are on a different scale from the LCC annotations (a continuous score from 0 to 1 versus a four-point scale), evaluation is not straightforward. Following Wang (2024), we reframe evaluation as a binary classification task, assigning negative labels to excerpts with a metaphoricity score of 0 and positive labels to those with a score of 1 or greater. Because we are interested in capturing even subtle metaphorical instances, this mapping aligns well with our goals. We report results at the threshold achieving the highest F1 score in Table 3. A full analysis with various thresholds is included in Appendix A.4.

We compare our prompting method against two alternative methods: a random baseline that assigns uniform random scores and a supervised RoBERTa classifier following Wang (2024), which fine-tunes the model on the LCC dataset using source span embeddings as input to the classification layer. For RoBERTa, we use 5-fold cross-validation and report micro-averaged results. To assess generalization across domains, we also evaluate RoBERTa in a leave-one-domain-out setting, fine-tuning on all but one target domain and testing on the held-out domain, and micro-averaged across all 32 domains.

The results are shown in Table 3. As expected, the supervised model achieves the highest macro F1, benefiting from in-domain training data. Our

Category	Canonical Groups
Person Groups	"immigrants", "non-immigrant u.s. citizens", "politicians", "law enforcement", "government", "other"
Place Groups	"U.S Country", "U.S. State", "U.S. City", "Non-U.S.", "unclear"
Thing Groups	"immigration", "money", "idea/opinion", "abstract concept", "other"

Table 4: Canonical target domains identified in hand-annotated set of tweets about immigration.

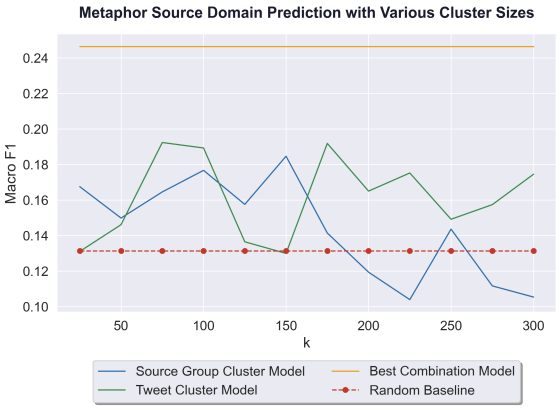

Figure 2: Performance of logistic regression models trained with various feature vectors. Best performance is achieved when metaphor source group cluster ($k = 50$) and tweet cluster ($k = 75$) information is combined (yellow).

prompting method with `Llama3.2` significantly over-classifies metaphor, achieving high recall at the cost of precision. However, our binary and score-based prompting methods with `Qwen3` perform competitively, and given that our approach requires no labeled data, we expect it to generalize more reliably to new domains and corpora.

To account for the possibility that the LCC dataset appeared in the pretraining data for `Qwen3` and inflated our results through data leakage, we additionally evaluate on a held-out set of examples with no publicly-available metaphor labels. We sample 100 `(source verb, target noun)` pairs from the unlabeled corpus of immigration-related tweets, balanced across LLM-generated metaphoricity scores, and have an author of the paper independently annotate each as "Metaphorical" or "Literal" using MIP (Group, 2007). We obtain an F1 score of 0.696 for this set, confirming the relatively high performance of the `Qwen3` score-based prompting method.

4.4 Quality of Domain Groups

We assess our target and source domain grouping methods using the immigration dataset. For the source domains, we measure the predictive signal of our induced groups for Mendelsohn and Budak (2025)'s source domain labels, accompanied by an intrusion test and qualitative analysis of the resulting groups. For the target domains, we conduct an annotation study to evaluate the quality of the group assignments.

4.4.1 Source Domain Groups

Predictive Signal Analysis To assess whether our induced source domain groups capture meaningful information, we frame the evaluation as a classification task—given a metaphorical tweet, predict the source domain labels of Mendelsohn and Budak (2025). To isolate the signal contributed by the groups themselves, we represent each tweet solely by its source domain group memberships, without incorporating any textual information, and train a logistic regression classifier on this representation, restricting to tweets with a metaphoricity score exceeding 0.3. To do this, we construct feature vectors of length k, where k is the number of clusters used to group LLM-generated metaphor explanations. All values of the feature vector are initialized to 0. If the tweet contains a metaphorical `(source verb, target noun)` pair that belongs to a given cluster i, the feature at index i is set to the metaphor salience score for the pair. If a single tweet has multiple pairs belonging to the same cluster, we take a mean of the metaphor scores. We note that our source domain groups, induced at the verb level from general metaphorical language, are not directly comparable to Mendelsohn and Budak (2025)'s tweet-level labels, which are assigned from a pre-defined set of dehumanization-focused domains. We therefore expect a performance ceiling, and instead focus on whether our induced groups carry meaningful predictive signal independent of the underlying text.

We compare against two baselines: a random baseline, and a *Tweet Cluster Model*, a logistic regression classifier using the same feature vector approach but clustering tweet context embeddings directly rather than metaphor explanations, serving as a metaphor-agnostic point of comparison. We also report the best performance obtained by combining the *Source Domain Cluster* and *the Tweet Cluster* feature vectors for all values of k.

Macro F1 scores for this predictive signal anal-

Source Domain	Metaphor Explanations
"physical object"	(1) *exhibited* is used metaphorically, framing *patriotism* as a tangible object that can be displayed.
	(2) *shows* is used metaphorically, framing *madness* as something that can be physically displayed.
	(3) *expose* is used metaphorically, framing the *political intentions* as the physical act of uncovering.
"money"	(1) *paying* is used metaphorically, framing the act of *allocating attention* as a financial transfer.
	(2) *fund* is used metaphorically, framing *policy support for immigration* as financial backing.
	(3) *cost* is used metaphorically, framing *losing lives* as a monetary expenditure.
"war"	(1) *won* is used metaphorically, framing a *political struggle* as a physical battle with a clear victor.
	(2) *fought* is used metaphorically, framing the *struggle for sovereignty* as physical combat.
	(3) *win* is used metaphorically, framing a *political or strategic struggle* as a physical battle to be won.

Table 5: Qualitative analysis of source group clusters from online immigration discourse.

ysis are shown in Figure 2. The results show that the models trained on the metaphor and tweet cluster feature vectors each outperform the random baseline across various values of k. Maximum performance is achieved by combining the two feature vectors, suggesting that the verb-level metaphor groups and text-based tweet groups capture complementary information for identifying Mendelsohn and Budak (2025)'s tweet-level source domains.

Intrusion Test We also evaluate the resulting source domain groups with an intrusion test. We evaluate the Metaphor Level clusters with $k = 50$ (see Figure 2). We construct triplets of LLM-generated metaphor explanations where two samples are from a given cluster and the third is from a different cluster. Annotators are asked to identify the intruder, i.e., the sample that does not fit with the others. We construct 100 triplets under three difficulty settings by ranking samples within each cluster by proximity to the centroid. In the "easy" setting (33 triplets), the two matching samples are drawn from the top 25% of the cluster, making them more semantically similar and easier to distinguish from the intruder. In the "medium" setting (33 triplets), they are drawn from the top 50%, and in the "hard" setting (34 triplets), from the top 75%. Each triplet is annotated by two annotators. We obtain a Krippendorff's α of 0.939, indicating high agreement and suggesting that the clusters are cohesive across difficulty levels.

Qualitative Analysis We inspect clusters for cohesiveness and identify source domains that are well documented CMT. Table 5 shows three examples: "physical object," "money," and "war" (Lakoff and Johnson, 2024).

4.4.2 Target Domain Groups

We identify 16 canonical target domains, shown in Table 4. To evaluate the quality of the target domain assignments, we recruit three students (same as Section 4.2) to annotate 100 randomly selected `(source noun, target group)` pairs with a metaphoricity score greater than 0.3. For each pair, annotators determine whether the example maps to the specified target noun and domain groups.

The annotators achieve a Krippendorff's α of 0.728 on noun group membership and 0.849 on target domain membership. `Qwen3` assigns noun groups with an accuracy of 0.913 and target domains with an accuracy of 0.717.

5 Case Study on Immigration Discourse

In this section, we outline our case study on immigration discourse. We use the larger split of $\sim 35,000$ immigration tweets, which include induced general policy frame, issue-specific frame, and narrative frame labels. These frames represent media frames, i.e., interpretive lenses used to emphasize particular aspects of an issue and shape how it is understood in public discourse. Our analysis is designed to ascertain whether the metaphor source domain clusters contribute signal for predicting frame labels. A detailed explanation for each frame can be found in Mendelsohn et al. (2021).

Predictive Signal We train and evaluate three logistic regression classifiers for frame label prediction: (1) *Source Group Cluster Model* (2) *Tweet Cluster Model* and (3) *Best Combination Model*. Features for these models are described in Section 4. We select k, the number of clusters for each model, based on the results shown in Figure 2,

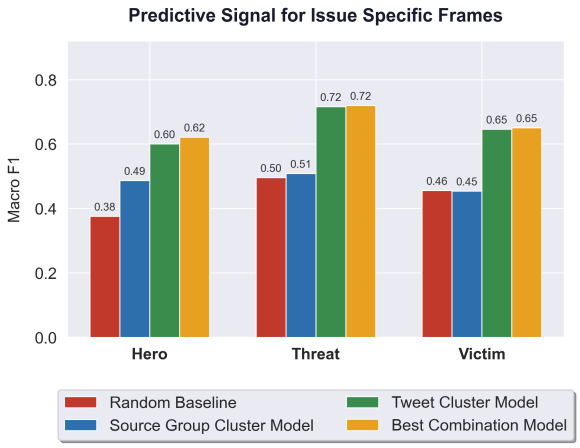

Figure 3: Performance of logistic regression models trained with various feature vectors to detect Issue Specific Frames in a dataset of tweets about immigration.

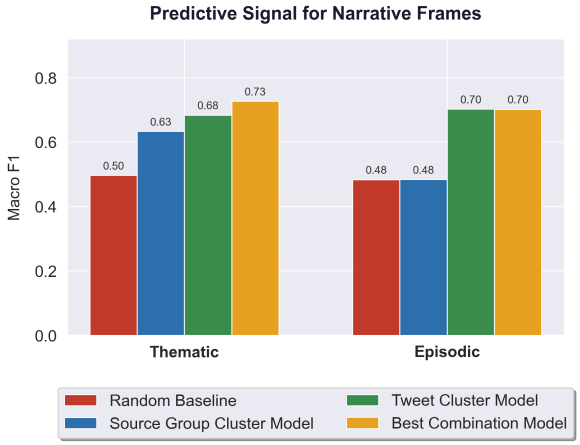

Figure 4: Performance of logistic regression models trained with various feature vectors to detect Narrative Frames in a dataset of tweets about immigration.

where maximum performance is achieved when metaphors are clustered with $k = 50$ and tweets are clustered with $k = 75$. Macro F1 results for predicting issue-specific frames are shown in Figure 3, narrative frames are shown in Figure 4, and general frames are shown in Figure 5.

We find that each of the three models typically outperforms the random baseline in predicting frame labels. An exception is that the *Source Group Cluster Model* fails to outperform the random baseline to predict "Episodic," "Security and Defense," and "Crime and Punishment" frames. The *Tweet Cluster Model* outperforms the *Source Group Cluster Model* for all frames. Consistently, the *Best Combination Model* achieves the best Macro F1, outperforming the *Tweet Cluster Model* in predicting the "Hero," "Thematic," "Health and Safety," "Policy Prescription and Evaluation," and "Crime and Punishment" frames. This shows that while our induced metaphors do not provide more signal for frame labels than the tweet context, they *enhance* it, suggesting that metaphors provide complementary information beyond the tweet text. These results are statistically significant at $p = 0.05$ after Nadeau-Bengio corrections, calculated with a paired t-test.

Qualitative Analysis We identify and describe source domain clusters with high feature importance for predicting a selection of frames in Table 6. This analysis shows relatively intuitive links between frames and clusters. For example, a cluster encompassing metaphorical uses of "build" is associated with the "Hero" frame, while metaphorical uses of "kill" are associated with both the "Health and Safety" and "Crime and Punishment" frames. However, this analysis also reveals some weaknesses in our clustering process. For example, cluster (14), associated with "make" is not cohesive. The physical act of creating something is applied to a wide range of abstract processes.

Qualitative analysis of Dehumanizing Metaphors We examine the resulting clusters for evidence of the dehumanizing metaphorical source domains widely studied in immigration discourse (Card et al., 2022; Mendelsohn and Budak, 2025). We find that metaphorical expressions which link to common dehumanizing source domains were not sorted cleanly into individual clusters. Instead, many of the dehumanizing metaphors congregated in the same "dehumanization" cluster, encompassing verbs such as "flooding" and "drowning" which maps to the "natural disaster domain" as well as "caged," which maps to the "animal" source domain. Prior work shows that these two source domains tend to be used by speakers with opposing stances on immigration issues, with the "natural disaster" domain favored by the right, and the "animal" domain favored by the left (Mendelsohn and Budak, 2025). Conflating them in a single cluster therefore obscures politically meaningful distinctions, motivating a more sophisticated grouping process for frame analysis in politically charged discourse.

6 Conclusion and Future Work

In this work, we developed an unsupervised method informed by CMT for identifying metaphorical expressions and grouping them to surface conceptual mappings that produce similar framings of

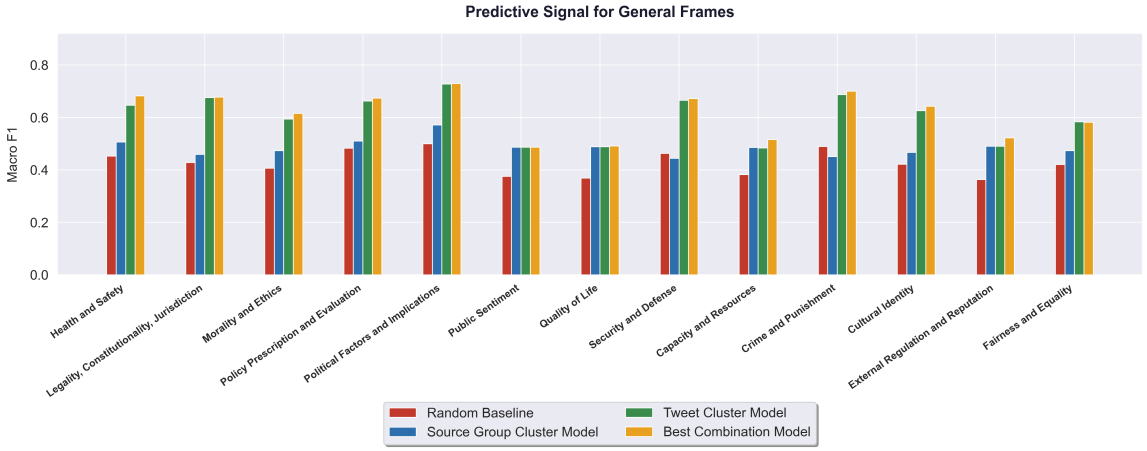

Figure 5: Performance of logistic regression models trained with various feature vectors to detect General Frames.

Frame	Important Cluster Descriptions
Hero	(34) *Vote, count, elect* and *cast* applied to contexts where literal voting is impossible. It is extended metaphorically to imply influence, endorsement, support, or removal.
	(24) *Build*, *create*, *construct*, and *manufacture* applied to abstract outcomes (e.g., nations, crises, policies) Physically bringing something into existence is mapped onto causing or establishing.
	(14) *Make* across a wide range of constructions applied to abstract outcomes. The physical act of making something is mapped onto a sprawling set of non-physical processes.
Health and Safety	(16) *Kill* and *murder* applied to people, groups, policies, institutions. Causing physical death is mapped onto the abstract processes of damaging, undermining, endangering, etc.
	(25) *Hit, attack, target, shoot, blast* (physical violence verbs) are mapped onto criticism, opposition, and strategic effort directed against people, policies, and institutions.
	(6) *Hurt* applied to countries, economies, communities, etc. Physical suffering is mapped onto being adversely affected by policies, economic forces, social conditions, or interpersonal actions.
Crime and Punishment	(16) described above
	(21) *Fix* applied to immigration systems, laws, crises, policies. Mending an object is mapped onto addressing, reforming, or resolving systemic dysfunction in laws, institutions, and social conditions.
	(45) *Break*, *cut*, *tear*, and *rip* applied to laws, systems, families, promises, countries. Destroying objects is mapped onto transgressing rules, dismantling relationships, and reducing resources.

Table 6: Descriptions of clusters important for predicting various frames.

an issue. Our framework achieves strong performance on unsupervised metaphorical expression detection, and the induced conceptual mappings are both quantitatively coherent (as measured by an expression-intrusion task) and qualitatively meaningful, capturing several well-documented source domains from the CMT literature. A case study on online immigration tweets demonstrates that the induced conceptual mappings provide predictive signal for frame labels and offer an interpretable way to study the relationship between conceptual metaphors and framing.

These results constitute a proof-of-concept for an automated conceptual mapping discovery framework. We identify three main directions for fully realizing this vision. First, we require a more thorough evaluation to conclude that our framework is domain-agnostic. Here we present a case study on immigration discourse, a domain where polarized metaphor usage is well-documented. In future efforts, we must determine if we can uncover interesting insights for other contentious domains where metaphor is understudied, such abortion or gun control. Second, we must develop and evaluate a method for mapping source words to distinct, named source concepts (rather than unnamed clusters) for more explicit conceptual mapping discovery. Third, for richer frame analysis, we must develop a method to induce Entman (1993)'s frame elements from the discovered conceptual mappings.

Limitations

This work has two main limitations. First, further evaluation is needed to determine how well our metaphorical phrase detection component generalizes to new domains and types of documents. In addition, we present only one case study on immigration discourse, which offers a limited view of the broader applicability of our approach.

Second, while we use a clustering approach to approximate source domain groups, we do not explicitly map metaphorical source words to distinct source concepts. Similarly, we do not explicitly explore the connection between the induced source and target concepts, which could provide additional insights and potentially guide improvements in metaphor induction.

Ethical Considerations

We recognize that using an LLM-based approach in a politically-charged domain may necessarily result in some biases and thus acknowledge a degree of uncertainty in reporting all results. We leave exploration of the specific biases of the `Qwen3` model in this area to important future work. Additionally, while one goal of this paper is to call attention to and condemn the use of dehumanizing metaphors, we recognize the potential for this line of work to reinforce such biases or for it to be utilized by bad actors.

Acknowledgments

This work was partially supported by a Research & Innovation Office (RIO) Seed Grant from the Office of the Provost and Executive Vice Chancellor for Academic Affairs, and RIO at the University of Colorado Boulder. Any opinions, findings, conclusions, or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the University or the Funding Offices.

This work utilized the Blanca high performance computing resources at the University of Colorado Boulder. Blanca is jointly funded by computing users and the University of Colorado Boulder.

References

Rodrigo Agerri, John Barnden, Mark Lee, and Alan Wallington. 2007. Metaphor, inference and domain independent mappings. In *RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING*, page 7.

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.

Jonathan Charteris-Black. 2016. *Fire Metaphors: Discourses of Awe and Authority*, 1 edition. Bloomsbury Publishing.

Lena Dankin, Kfir Bar, and Nachum Dershowitz. 2022. Can yes-no question-answering models be useful for few-shot metaphor detection? In *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*, pages 125–130, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Erik-Lân Do Dinh and Iryna Gurevych. 2016. Token-level metaphor detection using neural networks. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 28–33, San Diego, California. Association for Computational Linguistics.

Abhimanyu Dubey et al. 2024. The llama 3 herd of models. *ArXiv*, abs/2407.21783.

Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4):51–58.

Dan Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Jerome Feldman and Srinivas Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and language*, 89(2):385–392.

Matteo Fuoli, Weihang Huang, Jeannette Littlemore, Sarah Turner, and Ellen Wilding. 2025. Metaphor identification using large language models: A comparison of rag, prompt engineering, and fine-tuning. *Preprint*, arXiv:2509.24866.

Joseph Grady. 1997. Theories are buildings revisited. *Cognitive Linguistics*, 8(4):267–290.

Pragglejaz Group. 2007. Mip: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1):1–39.

Ilana Heintz, Ryan Gabbard, Mahesh Srivastava, Dave Barner, Donald Black, Majorie Friedman, and Ralph Weischedel. 2013. Automatic extraction of linguistic metaphors with LDA topic modeling. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 58–66, Atlanta, Georgia. Association for Computational Linguistics.

Rebecca M. M. Hicke and Ross Deans Kristensen-McLachlan. 2024. Science is exploration: Computational frontiers for conceptual metaphor theory. *Preprint*, arXiv:2410.08991.

Nicholas Ichien, Dušan Stamenković, and Keith J. Holyoak. 2024. Large language model displays emergent ability to interpret novel literary metaphors. *Preprint*, arXiv:2308.01497.

Zoltán Kövecses. 2000. The scope of metaphor. *Topics in English linguistics*, 30:79–92.

Zoltán Kövecses. 2017. *The Routledge handbook of metaphor and language*, chapter Conceptual metaphor theory. Routledge.

Zoltán Kövecses. 2020. *A Brief Outline of "Standard" Conceptual Metaphor Theory and Some Outstanding Issues*, page 1–21. Cambridge University Press.

G. Lakoff and M. Johnson. 2024. *Metaphors We Live By*. University of Chicago Press.

George Lakoff. 1993. The contemporary theory of metaphor.

George Lakoff. 2004. *Don't Think of an Elephant!: Know Your Values and Frame the Debate: the Essential Guide for Progressives*. Chelsea Green Publishing Company.

George Lakoff, Jane Espenson, and Alan Schwartz. 1991. Master metaphor list (technical report). *Cognitive Linguistics Group University of California, Berkeley*.

George Lakoff and Mark Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.

Yu Xi Li, Bo Peng, Yu-Yin Hsu, and Chu-Ren Huang. 2024. EmbodiedBERT: Cognitively informed metaphor detection incorporating sensorimotor information. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 16868–16876, Miami, Florida, USA. Association for Computational Linguistics.

James H Martin. 1990. *A computational model of metaphor interpretation*. Academic Press Professional, Inc.

James H Martin. 1994. Metabank: A knowledge-base of metaphoric language conventions. *Computational Intelligence*, 10(2):134–149.

Zachary J. Mason. 2004. CorMet: A computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.

Julia Mendelsohn and Ceren Budak. 2025. When people are floods: Analyzing dehumanizing metaphors in immigration discourse with large language models. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8079–8103, Vienna, Austria. Association for Computational Linguistics.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Julia Mendelsohn, Yulia Tsvetkov, and Dan Jurafsky. 2020. A framework for the computational linguistic analysis of dehumanization. *Frontiers in Artificial Intelligence*, 3:55.

George A. Miller. 1994. WordNet: A lexical database for English. In *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*.

Michael Mohler, David Bracewell, Marc Tomlinson, and David Hinote. 2013. Semantic signatures for example-based linguistic metaphor detection. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 27–35, Atlanta, Georgia. Association for Computational Linguistics.

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).

Srinivas Narayanan. 1999. Moving right along: A computational model of metaphoric reasoning about events. *Aaai/iaai*, 121127.

Paolo Pedinotti, Eliana Di Palma, Ludovica Cerini, and Alessandro Lenci. 2021. A howling success or a working sea? testing what BERT knows about metaphors. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 192–204, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Miriam R L Petruck and Ellen K Dodge. 2016. MetaNet: Repository, identification system, and applications. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, Berlin, Germany. Association for Computational Linguistics.

E. Puraivan, I. Renau, and N. Riquelme. 2024. Metaphor identification and interpretation in corpora with ChatGPT. *SN Computer Science*, 5:976.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Elisa Sanchez-Bayona and Rodrigo Agerri. 2025. Metaphor and large language models: When surface features matter more than deep understanding. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 17462–17477, Vienna, Austria. Association for Computational Linguistics.

Ekaterina Shutova. 2010. Models of metaphor in NLP. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 688–697, Uppsala, Sweden. Association for Computational Linguistics.

Ekaterina Shutova and Lin Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 978–988, Atlanta, Georgia. Association for Computational Linguistics.

Ekaterina Shutova, Lin Sun, Elkin Darío Gutiérrez, Patricia Lichtenstein, and Srini Narayanan. 2017. Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning. *Computational Linguistics*, 43(1):71–123.

Ekaterina Shutova, Lin Sun, and Anna Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1002–1010, Beijing, China. Coling 2010 Organizing Committee.

Krishnkant Swarnkar and Anil Kumar Singh. 2018. DiLSTM contrast : A deep neural network for metaphor detection. In *Proceedings of the Workshop on Figurative Language Processing*, pages 115–120, New Orleans, Louisiana. Association for Computational Linguistics.

Qwen Team. 2025. Qwen3 technical report. *Preprint*, arXiv:2505.09388.

Yuan Tian, Nan Xu, and Wenji Mao. 2024. A theory guided scaffolding instruction framework for LLM-enabled metaphor reasoning. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 7738–7755, Mexico City, Mexico. Association for Computational Linguistics.

Xiaoyu Tong, Rochelle Choenni, Martha Lewis, and Ekaterina Shutova. 2024. Metaphor understanding challenge dataset for LLMs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3517–3536, Bangkok, Thailand. Association for Computational Linguistics.

Yunxiao Wang. 2024. Metaphorical framing of refugees, asylum seekers and immigrants in UKs left and right-wing media. In *Proceedings of the 8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL 2024)*, pages 18–27, St. Julians, Malta. Association for Computational Linguistics.

Linyi Yang, Yaoxian Song, Xuan Ren, Chenyang Lyu, Yidong Wang, Jingming Zhuo, Lingqiao Liu, Jindong Wang, Jennifer Foster, and Yue Zhang. 2023. Out-of-distribution generalization in natural language processing: Past, present, and future. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4533–4559, Singapore. Association for Computational Linguistics.

Jintao Zhang, Guoliang Li, and Jinyang Su. 2025. Sage: A framework of precise retrieval for rag. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*, pages 1388–1401.

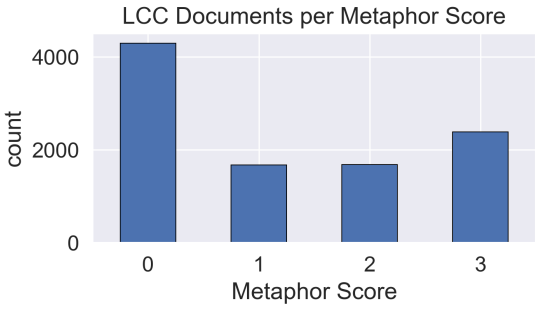

Figure 6: Distribution of metaphor scores in LCC split used to benchmark our metaphor classification method. When we consider all documents with a score of 1 or greater to be metaphorical,

A Appendix

A.1 Additional Metaphor Induction Details

In Table 7 we present the five constructional patterns used to filter `(verb, noun)` pairs which can grammatically induce a metaphor. In Table 7 we show a sample of LLM metaphor salience score explanations, which are clustered to group metaphors with similar source domains.

A.2 Additional Dataset Details

Table 8 and Figure 6 shows the distribution of hand-annotated metaphoricity scores for the 10,018 documents from the LCC dataset used to evaluate our metaphor classification method. Table 9 shows the hand-annotated target domains for this set.

A.3 Additional Binary Metaphor Classification Results

Table 10 shows a detailed classification report of our implementation of the supervised metaphor classification component presented in (Wang, 2024). Our classifier achieves an F1 score of 0.809 on all positive classes (metaphor score greater than 0), resulting in a macro F1 score of 0.810 when the task is viewed from the binary classification perspective. These results are very similar to those achieved by (Wang, 2024), who presented an F1

Construction Pattern	Example
S_VERB-*dobj*-T_NOUN	"I gave you that idea."
T_NOUN-*nsubj*-S_VERB	"Ideas travel quickly."
S_VERB-*agent*-ADP-*pobj*-T_NOUN	"That idea gnaws at my mind."
S_VERB-*amod*-T_NOUN	"That idea sings."
T_NOUN-*nsubjpass*-S_VERB	"My idea was shot down."

Table 7: Constructional patterns used to filter SDPs which may invoke a metaphor. Italics in the "Construction Pattern" column represent relation types between tokens. Source verbs and target nouns are underlined in the "Example" column.

LCC Score	Count
0	4,285
1	1,671
2	1,677
3	2,385
1 $\leq$	5,733
total	10,018

Table 8: Distribution of metaphor scores for the portion of the LCC dataset used for evaluation.

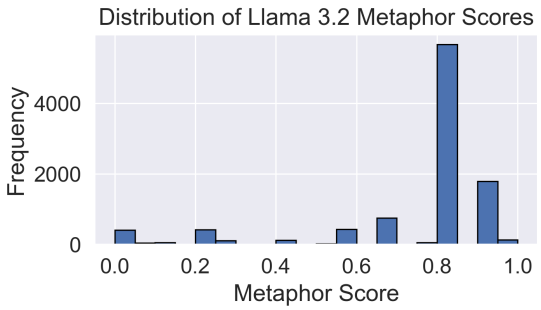

Figure 7: Llama 3.2-generated metaphor salience scores for (`source verb, target noun`) pairs extracted from the LCC dataset.

score of 0.86 on samples with a metaphor score of 0 and 0.83 on those with a score greater than 0.

A.4 Additional Metaphoricity Threshold Analysis

Tables 7 and 8 show the distribution metaphor salience scores for (`source verb, target noun`) pairs in the LCC dataset. Table 11 shows the performance of the scores on the LCC dataset for metaphor classification at various thresholds.

LCC Target Domain	Count
guns	2013
mental concepts	1577
government	1025
democracy	673
elections	613
bureaucracy	603
poverty	581
taxation	498
wealth	430
religion	409
money	313
disease	231
migration	152
intellectual property	125
taxpayers	104
taxes	99
islamic	86
terrorism	85
gun debate groups	79
gun rights	78
politicians	78
marriage	45
drug trafficking	40
welfare	23
debt	16
climate change	15
abortion	13
demographics	11
control of guns	3
total	10018

Table 9: Hand-annotated target domains for all excerpts in the English LCC dataset used for evaluation. The dataset spans a wide variety of topics, making it an optimal resource for evaluating computational methods for general metaphor detection.

	precision	recall	f1-score	support
0	0.778	0.848	0.812	2403
1	0.438	0.347	0.387	654
2	0.399	0.403	0.401	705
3	0.718	0.668	0.692	1218
accuracy	0.675	0.675	0.675	0.675
macro avg	0.583	0.566	0.573	4980
weighted avg	0.665	0.675	0.669	4980

Table 10: RoBERTa classifier performance trained and tested on the complete LCC dataset. Here the task is formulated as a multi-class classication task to predict metaphor scores 0-3.

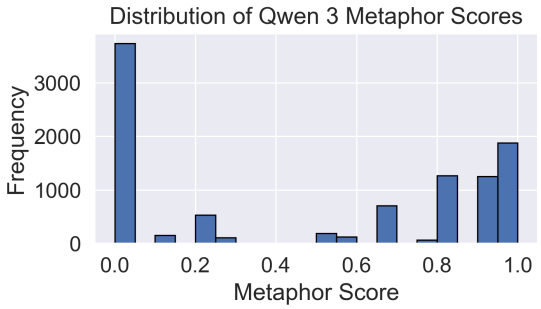

Figure 8: Qwen 3-generated metaphor salience scores for (`source verb, target noun`) pairs extracted from the LCC dataset.

Threshold	**0.1**	**0.2**	**0.3**	**0.4**	0.5	0.6	0.7	0.8	0.9	1
Random Baseline	**0.699**	0.663	0.623	0.579	0.525	0.462	0.386	0.291	0.158	0
Llama 3.2	0.732	**0.733**	0.732	0.732	0.731	0.731	0.725	0.705	0.333	0.027
Qwen 3	0.787	0.790	**0.792**	0.791	0.791	0.786	0.778	0.733	0.613	0.426

Table 11: Macro F1 performance of our method on the LCC dataset at different score thresholds. LCC documents with a metaphorical score of 1 or higher are considered metaphorical. Documents with an LLM score above the threshold are predicted to be metaphorical.

System Prompt

You are an annotator who is developing a dataset for measuring metaphors. Your response should be in JSON format with the key 'metaphor_salience_score' and a value between 0 and 1 that indicates how salient the metaphor invoked by the specified verb is. Additionally, in the JSON, you should indicate with the key 'explanation' the justifications for your decision.

Desired JSON: {'metaphor_salience_score': float, 'explanation': 'your reasoning for the answer'}.

Do not generate anything else.

User Prompt

Analyze the use of the specified verb in the sentence provided. Focus only on determining whether this verb, in its specific use within the sentence, is used literally, that is, describing the physical action, or whether it is used metaphorically, where the use of the verb transcends its original meaning without referring directly to a physical action, such as, for example, giving some kind of personification or animalization of an object. It is important to distinguish the specific lexical analysis of the verb from any broader metaphorical interpretation that may arise from comparisons or conceptual equivalences present in the sentence.

Please give a score between 0 and 1 (inclusive) that indicates how salient the metaphoricity of the specified verb is. 0 indicates that the verb is very obviously used literally and 1 indicates that the verb is very obviously used metaphorically. A score of 0.5 indicates that the metaphor would only be noticed by half of human readers. Please provide your evaluation focusing solely on the specified verb.

Prompt A.1: Metaphor salience score extraction.

System Prompt

You are a grammar expert building a dataset for noun classification. Your task is to determine whether the target noun is a person (or group of people), place, or thing.

Desired JSON: {'noun_classification': '[person/place/thing]'}.

Do not generate anything else.

User Prompt

Is the the target noun specified below a person (or group of people), place or thing? You may use the context sentence to help make your decision, but please provide your answer focusing solely on the specified noun.

TARGET WORD: <target word to classify>
SENTENCE: <context sentence of target word>

Prompt A.2: Assigning person, place, or thing classifications to target words.

System Prompt

You are an annotator who is developing a dataset for analyzing metaphors. Your task is to characterize how metaphors are being used to characterize a particular person or group of people. For each submission, you will identify the target group that matches the provided metaphor target word.

ANNOTATION GUIDELINES
- Map the identified nouns to the most appropriate group.
- If no group from the provided list adequately represents the target, assign it to 'Other'. This includes nouns that are tangentially related but don't fit well into any specific predefined category, as well as nouns from completely different domains or contexts.

Desired JSON: {'target_group': ['Identified character group from predefined list']}.

Do not generate anything else.

User Prompt

Please provide an evaluation focusing solely on the metaphor target noun.

DOMAIN: <corpus domain>
GROUPS: <identified target domains>
TARGET WORD: <target word to classify>
SENTENCE: <context sentence of target word>

Prompt A.3: Assigning target domains to target words.