Not all ANIMALs are equal: metaphorical framing through source domains and semantic frames

Anonymous ACL submission

Abstract

Metaphors are powerful framing devices, yet their source domains alone do not fully explain the specific associations they evoke. We argue that the interplay between source domains and semantic frames determines how metaphors shape understanding of complex issues, and present a computational framework that allows to derive salient discourse metaphors through their source domains and semantic frames. Applying this framework to climate change news, we uncover not only well-known source domains but also reveal nuanced frame-level associations that distinguish how the issue is portrayed. In analyzing immigration discourse across political ideologies, we demonstrate that liberals and conservatives systematically employ different semantic frames within the same source domains, with conservatives favoring frames emphasizing uncontrollability and liberals choosing neutral or more ”victimizing" semantic frames. Our work bridges conceptual metaphor theory and linguistics, providing the first NLP approach for discovery of discourse metaphors and fine-grained analysis of differences in metaphorical framing.[1]

1 Introduction

Metaphors help us understand and explain our world by transferring what we know about physical, tangible objects to more abstract, hard-to-define concepts and notions, or, as Lakoff and Johnson (2008) succinctly put it, “understanding and experiencing one kind of thing in terms of another”. In particular, Lakoff and Johnson (2008) showed that **target** (more abstract) concepts are understood in terms of more concrete, physical **source** domains, such as LOVE IS BATTLE or THEORIES ARE BUILDINGS. Following their seminal work, *conceptual* metaphor theory (CMT) became a productive method to generalize from specific metaphorical expressions and arrive at a more abstract analysis of associations carried over from the source domain. In NLP, work on metaphorical understanding also focuses on mapping metaphors to their source domains (Shutova and Teufel, 2010; Mohler et al., 2016; Mendelsohn and Budak, 2025).

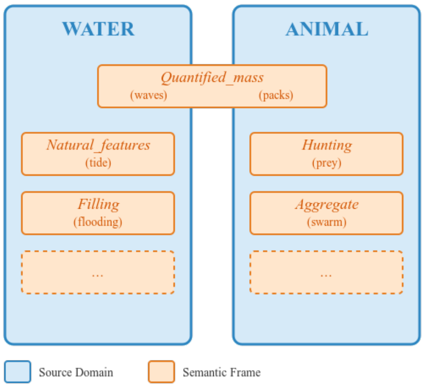

Figure 1: Interaction between semantic frames and domains

However, the source domain itself does not fully explain which associations are carried over. Consider the following metaphors, both deriving from the source domain of WATER, commonly used in immigration discourse[2]:

(1) Illegal aliens continue to *flood* into our country, ruining our economy.

(2) Maybe a high *tide* raises all boats? *Waves* of immigrants have always enriched us.

The first example compares immigration to *filling* with water, which must be controlled and stopped; the second treats it as a *mass of water* which, however, does not pose a risk of "overflowing" and can be considered a natural *feature of*

[1] Code and data are available at `https://anonymous.4open.science/r/ConceptFrameMet-50F8/`. Pre-trained models will be released upon acceptance.

[2] Simplified examples from Mendelsohn and Budak (2025)’s dataset.

landscape, thus instilling a more positive, accepting attitude towards immigration.

Such associations have been explained by linguistic, constructionalist theories of metaphors (Sullivan, 2025, 2013), positing that the associations derived from the source domain are related to the *semantic frame* of a lexical item used in the metaphor (Fillmore and Baker, 2001). In our example, *flood* has a semantic frame of *Filling* which emphasizes the movement of water and the negative result of such movement. On the other hand, *wave* and *tide* have semantic frames of *Quantified_mass* and *Natural_features* with more neutral associations. Semantic frames are generic structures that can be applied to many source domains. For example, the *Quantified_mass* semantic frame can also be applied to the ANIMAL source domain (*packs of immigrants*) or PRESSURE/PHYSICAL_BURDEN source domain (*loads of illegals*). Thus, it is the interplay of the source domain and semantic frame that uniquely defines the associations of the metaphor, where the source domain points to a cluster of associations, and the semantic frame allows to pick out specific ones (Figure 1).

This nuanced understanding of associations is particularly important when metaphors are being used as *framing devices* to emphasize and highlight a particular aspect of an issue or a debate, i.e. to frame it (Entman, 1993).[3] These discourse metaphors or "relatively stable metaphorical projections that function as key framing devices within a particular discourse over a certain period of time" (Zinken et al., 2008; Nerlich and Jaspal, 2012) are the focus of our study. Specifically, we show that differences in metaphor use between political ideologies — which has not been fully explained in prior work (Mendelsohn, 2024; El Refaie, 2001) — can be effectively analyzed through consistent, prevalent choice of semantic frames that funnel particular associations from the source domain.

To support our analysis of discourse metaphors, we implement a framework that consists of two components: a pre-trained language model that detects metaphors and predicts their semantic frame and source domain, and a statistical module that uses log-likelihood ratio estimation (Rayson and Garside, 2000) to highlight salient source domains and semantic frames of metaphors within a particular discourse compared to a control corpus.

We apply our framework to analyze metaphorical framing in two scenarios. First, we discover discourse metaphors used to frame *climate change* and not only discover well-known source domains of climate change metaphors (such as WAR), but also reveal more nuanced semantic frame associations consistently chosen within the source domains. Second, we analyze the source domains of WATER and ANIMAL in metaphors in immigration discourse, comparing their use by liberals and conservatives (Mendelsohn and Budak, 2025), and show that these domains are heterogeneous, and that tweets with different political leanings routinely different semantic frames of metaphors within the same domains to convey particular emotions towards immigrants.

In sum, we are the first to connect source domains and semantic frames in an NLP approach; release a manually labeled corpus in the climate domain; and apply our framework across two domains showing it can be used to *discover* discourse metaphors within a particular topic, and *highlight nuanced differences* in superficially similar metaphor usage.

2 Related work

Metaphor analysis in NLP Metaphors have attracted significant attention, with an extensive body of work focusing on annotating metaphors and creating corpora (Group, 2007; Steen et al., 2010; Boisson et al., 2025), automatically detecting them in text (Wang et al., 2025; Zhang and Liu, 2022; Uduehi and Bunescu, 2024; Tian et al., 2024; Reimann and Scheffler, 2025), understanding (Tong et al., 2024; Ye et al., 2025), generation (Joseph et al., 2023; Chakrabarty et al., 2021; Veale, 2016), and interpretation, such as the ability to handle them in inference and question answering tasks (Liu et al., 2022; Sanchez-Bayona and Agerri, 2025) (see Ge et al. (2023); Rai and Chakraverty (2020) for an extensive survey).

Concept metaphor theory (CMT) (Lakoff and Johnson, 2008) has been widely adopted as a tool to improve detection, explanation, and generation of metaphors (Jones, 1992; Stowe et al., 2021a; Ge et al., 2022; Mao et al., 2023; Tian et al., 2025). Some works also use semantic frames as a proxy,

[3]The term *frame* refers to (at least) two distinct concepts – semantic frames as encoded in FrameNet (Fillmore and Baker, 2001), and media frames, i.e. consistent emphasis of particular aspects of an issue to evoke specific associations in reader's mind (Entman, 1993); see Otmakhova et al. (2024); Sullivan (2023) for the discussion of their relation to each other and distinctions. To avoid confusion, we use terms *semantic frames* and *media frames*.

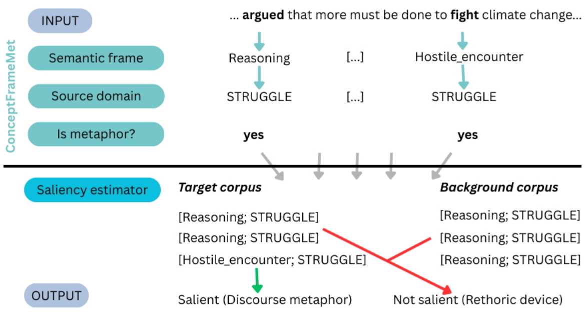

Figure 2: Framework overview showing semantic frames and source domains for the input metaphors (**bold**).

or substitute, for conceptual mappings, in order to detect (Li et al., 2023), or generate (Stowe et al., 2021b) metaphors. We are the first to leverage the interaction between the two frameworks.

According to the constructionist view of metaphors, their conceptual understanding is bound by linguistic constructs (Sullivan, 2013, 2025). Rosen (2018) use construction grammar cues such as argument structure to improve prediction of source domains of metaphors, Jang et al. (2017) collect syntactical patterns associated with a particular source domain for improved metaphor detection, while several studies encode semantic roles or domains (*agent*, *physical_affliction*) (Gordon et al., 2015; Dodge et al., 2015), establishing schema that consists of multiple lexical instantiations. However, with the exception of a small descriptive study which looks at the differences in semantic frames used for the same CMT metaphor across two languages (Gamonal, 2022), we are the first to use FrameNet-style semantic frames to contrast specific metaphor instantiations and pinpoint the differences in associations carried through them.

Metaphors as framing devices While metaphors are well-known framing devices (Landau et al., 2009; Boeynaems et al., 2017; Semino et al., 2018; Brugman et al., 2019), the link between metaphors and framing is still under-explored in NLP. Several studies approximated framing effects of metaphors through other variables: Prabhakaran et al. (2021) analyzed effects of metaphor usage on reader engagement, while Baleato Rodríguez et al. (2023) modeled metaphors to improve propaganda detection.

A number of works focus on metaphors as framing devices by analysing their source domains in a particular topic or discourse (Chen et al., 2022; Sengupta et al., 2022; Guan and Zeng, 2024; Li et al., 2024; Meng et al., 2025) or by statistically comparing source domains across political leanings (Sengupta et al., 2024; Wang, 2024; Mendelsohn and Budak, 2025). None of these studies, however, explain why metaphors from the same source domain are used across polar opposite ideologies (Mendelsohn and Budak, 2025), or considered finer distinctions within the source domain. We bridge this gap by adopting semantic frames to extract particular associations from source domains, and saliency as a tool to discover prevalent and distinctive combinations of source domains and semantic frames.

3 Methods

We introduce our two-part framework: a model that combines metaphor detection with predicting their source domain and underlying semantic frames (Section 3.1), and the log-likelihood ratio method used to estimate the saliency of source domains and semantic frames in the corpus (Section 3.2). Figure 2 shows an overview of the framework.

3.1 Predicting metaphors, source domains and semantic frames

In this section we describe ConceptFrameMet – a RoBERTa (Liu et al., 2019) model for detecting metaphors and predicting their source domain and semantic frames (Figure 2, top). The model relies on three components – Semantic frame classifier (Section 3.1.1), Source domain classifier (Section 3.1.2), and metaphor classifier (Section 3.1.3).

3.1.1 Semantic frame classifier

Data. We use FrameNet 1.7 (Ruppenhofer et al.) which contains sentences with semantic frames annotations over 797 classes, with the train, dev, and test splits of 19391/2272/6714 samples (Swayamdipta et al., 2017)).

Models. We fine-tune a RoBERTa-base model (Liu et al., 2019) to output a probability distribution over 797 semantic frames for a particular word in a sentence (details in Section A.1.1). We compare two input configurations, with the target word masked out (*immigrants MASK the cities*) and with the target word repeated separately from the context sentence (*SEP immigrants flood the cities SEP flood SEP*). We also prompt two LLMs – Gemini 2.5 and Claude Sonnet 4.0, which we prompt to choose one of 797 classes (details in Section A.2.1).

	Acc	micro-F1	macro-F1
RoBERTa MASK	0.806	0.806	0.053
RoBERTa SEP	**0.861**	**0.866**	**0.648**
Gemini 2.5	0.508	0.508	0.430
Claude Sonnet 4.0	0.736	0.736	0.600
An et al. (2023)	0.899	–	–
Devasier et al. (2024)	–	0.887	–

Table 1: Semantic frame prediction performance on FN1.7 test across three runs (sd < 0.05).

	Acc	P	R	F1
RoBERTa MASK	0.307	0.203	0.184	0.182
RoBERTa SEP	0.833	0.745	0.742	0.740
Frames_CONCAT	0.837	0.759	**0.758**	0.754
Frames_ATTN	**0.838**	**0.764**	0.757	**0.756**
Gemini 2.5	0.528	0.419	0.356	0.345
Claude Sonnet 4.0	0.528	0.517	0.452	0.445

Table 2: Source domain prediction performance on the LCC dataset across three runs (sd < 0.05). All metrics are macro-averaged.

We also include results reported on the same test set for the same task in two recent papers (An et al., 2023; Devasier et al., 2024), which, however, are knowledge-heavy requiring substantial data-augmentation and are thus unsuitable to deploy for our large-scale topic-specific analyses.

Results. As shown in Table 1, our best performing model (RoBERTa SEP) performs comparatively to the knowledge-heavy models in prior work. The LLMs significantly under-perform our fine-tuned RoBERTa models. Accordingly, we use the SEP-style model in all further experiments, which is light-weight and scalable to large data sets.

3.1.2 Source domain classifier

Data. We use the LCC Metaphor Dataset (large) (Mohler et al., 2016), the largest English dataset annotated for source domains, and pre-process it as described in A.1.2. The resulting dataset has 99 source domains, and is randomly sampled with stratification into train, dev, and test of 11704/2509/2509 samples.

Models. We fine-tune a RoBERTA-base model to predict one of 99 source domains (details in A.1.2). Similar to Section 3.1.1, we compare the masked input style with the inputs where the metaphor and the context are separated by SEP. Moreover, since some of the source domain classes are semantically close and easily confused (Shutova and Teufel, 2010; Mohler et al., 2016), we hypothesize that including the semantic frame of the metaphor will help the model to differentiate between domains. To test this, we pass the predicted probability distribution of semantic frames as a frozen feature vector to the model in two ways: by either concatenating it with the RoBERTa-encoded input vector (Frames_CONCAT), or by applying attention over semantic frames (together with the frozen vector as residuals) to highlight those semantic frames that are important for the source domain prediction (Frames_ATTN). We also compare the fine-tuned models with zero-shot Gemini 2.5 and Claude Sonnet 4.0, where the prompt includes the list of 99 classes (Section A.2.2).

Results. Table 2 confirms that adding semantic frame information to the input boosts source domain prediction performance, with the attention-based model performing best overall. In particular, they achieve improvement of 20 points of macro F1 on underrepresented classes (<10 instances in training data; see Figure 5 in the appendix). Most fine-tuned RoBERTa models again outperform zero-shot LLMs. We use the Frames_ATTN model in all further analyses.

3.1.3 Metaphor prediction

Data. We use VUA-18, the largest available metaphor prediction dataset (Leong et al., 2018) to train and test our models, and evaluate generalizability through zero-shot testing on two smaller benchmarks – MOH-X (Mohammad et al., 2016) and TroFi (Birke and Sarkar, 2006) (see Section A.1.3 for data statistics).

Models. We integrate our semantic frame and source domain classifiers with a metaphor prediction model. For the latter, we choose MelBert (Choi et al., 2021), which is the strongest pre-trained model among "knowledge-lean" models that do not require augmenting data with additional features. MelBert is a RoBERTa model which captures the inconsistency between the contextual meaning of the metaphor and its literal meaning through contrasting embeddings of the word in context and in isolation. We posit that the encoding of a word in isolation still mixes representations of literal and metaphorical usage; for example, the embedding of the isolated word *flooded* will likely reflect both the senses of *the river flooded the city* and *immigrants flooded the city*, i.e. still be somewhat close to a metaphorical context embedding. We thus replace the word embedding with that of the predicted source domain (WATER), amplifying the

	VUA-18	TroFi	MOH-X
Random baseline	0.222	0.466	0.486
MelBert	**0.782**	0.631	0.806
FrameBert	0.766	0.620	0.780
Gemini 2.5	0.341	0.633	**0.849**
Claude Sonnet 4.0	0.330	**0.651**	0.822
ConceptFrameMet (ours)	<u>0.767</u>	<u>0.634</u>	0.814

Table 3: Binary F1 (positive class) performance on metaphor prediction for models fine-tuned on VUA-18 (train). We show results on VUA-18 (test) and two other test sets. LLM results are zero-shot for all three datasets. Bold=best; underline=second best.

difference between contexts in the metaphorical case (*immigrants* and *water*), and minimizing the gap in literal cases (*river* and *water*).

To account for source domain prediction noise, instead of replacing a word with its predicted source domain, we blend their embeddings, using the confidence score of the source domain prediction (α) as a weight:

$$\mathbf{e}_{\text{blended}} = \alpha \cdot \mathbf{e}_{\text{source\_domain}} + (1 - \alpha) \cdot \mathbf{e}_{\text{word}}.$$

We compare the resulting model with FrameBert (Li et al., 2023), a modified version of MelBert and Claude Sonnet 4.0 and Gemini 2.5 which are prompted to identify all metaphors in a given text (Section A.2.3).

Results. Results in Table 3 show that MelBert performs best in-domain (i.e., when tested on the test portion of its training data set) but fails to generalize to other data sets. LLMs perform best on the smaller test sets. ConcepFrameMet (our method) shows the best overall performance, only slightly underperforming the best method across all three test sets, while being substantially more resource efficient than the LLMs. The LLMs tend to overpredict metaphors, so they struggle on VUA-18 where the rate of metaphors is low (15%). However, the results on VUA-18 are indicative of practical usage scenarios where one needs to distinguish metaphors from mostly literal words. This, together with low performance of LLMs on semantic frame and source domain prediction, motivates us to use our light-weight, integrated PLM model for metaphor detection and analysis.

3.2 Detecting discourse metaphors

To distinguish between *discourse metaphors* — which carry stable, prevalent associations that allow to resolve an ambivalence of interpretation of some issue (Scheufele and Scheufele, 2010) — and metaphors that are merely figures of speech, we define two principles. First, the target domain of discourse metaphors must be a polarizing issue. Second, discourse metaphors are salient in a given discourse (Zinken et al., 2008). We operationalize *saliency* as the log-likelihood ratio (Rayson and Garside, 2000), which is a common technique to compare distributions of items (such as words or, in our case, semantic frames and source domains) between two corpora. In particular, given two corpora C_1 and C_2, the log-likelihood ratio identifies candidate metaphors that are significantly over-represented in one of the corpora, i.e., reject the null hypothesis H_0 that they are represented equally in both corpora (C_1 and C_2).

$$-2 \ln \lambda = -2[\ell(\theta_0) - \ell(\hat{\theta})] = 2 \sum O_i \ln \left(\frac{O_i}{E_i} \right)$$

where λ = likelihood ratio, $\ell(\theta_0)$ = log-likelihood under H_0, $\ell(\hat{\theta})$ = maximum log-likelihood (from (C_1, C_2)), O_i = observed frequency, E_i = expected frequency under H_0.

Source domains and semantic frames with log-likelihood values that reject H_0 with $p > 0.05$ are identified and their relative frequency is compared to decide which of the corpora (C_1 or C_2) they are more strongly associated with.

4 Metaphorical framing through source domains and semantic frames

We now apply our framework explained in Section 3 to (1) *discover* discourse metaphors in texts about a particular topic by comparing metaphors in a topic-specific vs. a general back ground corpus (Section 4.1); and *contrast* discourse metaphors in texts associated with polar opposite political leanings within a single topic (Section 4.2).

4.1 Metaphorical framing in climate change news

Existing methods for metaphorical framing analysis relies on manual discovery of prevalent source domains in a particular discourse, while NLP approaches require such knowledge of source domains a priori. Here, we apply our framework to automatically identify salient source domains from a large collection of news on climate change, and further analyze them through salient semantic frames.

	Accuracy	F1
Metaphor prediction	0.955	0.976
Source domain prediction	0.935	0.728

Table 4: Performance of ConceptFrameMet on the climate change metaphors dataset. We report binary F1 for the metaphor prediction task, and macro-F1 for the source domain prediction task.

4.1.1 Data and methods

We use a corpus of New York Times articles published between 1986 and 2020 (Fast and Horvitz, 2017; Mendelsohn et al., 2020). We extract all paragraphs which mention "climate change" or "global warming", and then additionally filter them using a ClimateBERT pre-trained model (Webersinke et al., 2021) which was fine-tuned to detect climate-related texts (Bingler et al., 2023). This results in 47K paragraphs (distribution over years in Section A.4). For saliency analysis, we also collect a generic corpus which contains the same number of paragraphs randomly sampled from the same underlying NYT data set ensuring that the paragraphs do not contain our climate keywords.

Next, we use ConceptFrameMet (Section 3.1) to detect metaphors, their semantic frames and source domains, in both the climate and generic corpus. To ensure that the identified metaphors refer to the topic of interest, we retain only sentences that contain terms "climate" or "warming". Finally, we calculate log-likelihood ratios for source domains, and then for semantic frames within each source domain, as described in Section 3.2.

4.1.2 Evaluation

The process above resulted in a corpus of 6,859 sentences that contain metaphors with source domains and semantic frames salient in the climate change corpus ($p < 0.05$). To evaluate the performance of ConceptFrameMet on that corpus, we randomly select a sample of 400 sentences that covers all combinations of salient source domains and semantic frames within them. We presented crowd workers on Prolific with sentences containing highlighted metaphors and asked them to identify up to three source domains from a list of 5 options, comprising the top two domains predicted by ConceptFrameMet and three random distractors plus options OTHER DOMAIN and NO METAPHOR. See Section A.8 for more details including quality control, compensation and instructions We collected four annotators per sample and achieve a reliable averaged pair-wise annotator agreement of 69% , and use majority voting (with adjudication by one author this paper with a degree in linguistics) to determine the final source domain label. We finally compute model performance against human labels, judging a prediction as correct if the human majority vote corresponds to the top-1 model prediction (Table 4), showing that ConceptFrameMet generalizes well to our new domain.[4]

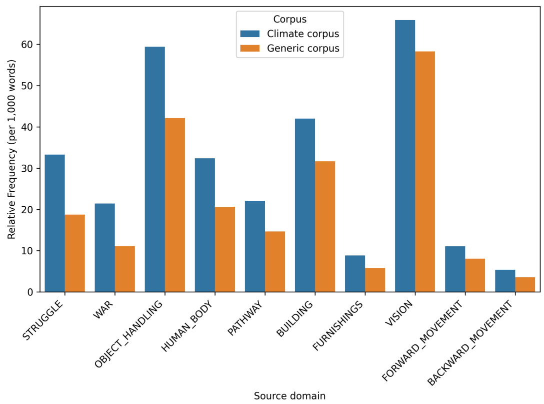

Figure 3: Relative frequency of source domains in climate and generic corpora. The source domains are sorted left to right in order of diminishing saliency

4.1.3 Results

Source domains Figure 3 presents the 10 most salient source domains in the climate corpus, ordered by log-likelihood ratio, from left to right, while the y-axis indicates frequency, noting that saliency highlights *differences in frequency* of source domains between corpora rather than the *absolute frequency* of a source domain within a corpus. Thus, a frequent source domain (such as VISION) can be ranked lower than a less frequent domain (STRUGGLE), and the most frequent source domain in the climate corpus (MOVEMENT) is not among the most salient ones. Importantly, the discovered source domains are well-aligned with the ones reported in theoretical critical discourse analysis studies. We present examples of metaphors from each of the top 10 source domains, together with the theoretical studies, in Section A.5.

Semantic frames Even within a salient source domain, some metaphors can be used as a figure of speech rather than a discourse (framing) metaphor.

[4]Note that the metaphor prediction F1 is inflated since all samples were predicted as containing metaphors by the model; in this we follow the process and limitations of the metaphor corpus creation suggested by (Mohler et al., 2016)

	Climate corpus		Generic corpus	
	Semantic frame	Metaphors	Semantic frame	Metaphors
STRUGGLE	*Hostile_encounter* *Topic* *Relation*	fight, confront address at odds	*Difficulty* *Resolve_problem* *Cause_to_end*	challenges, hard settle ended
WAR	*Hostile_encounter* *Boundary* *Judgment_communication*	battle, waged war on the front lines assailed, crusaded	*Change_of_leadership* *Invading* *Aiming*	rebel, revolution invasion, intervene target
OBJECT HANDLING	*Attempt_suasion* *Taking_sides* *Intentionally_act*	push (for), press take (a stand), handle take (action, steps)	*Getting* *Intercepting* *Entity*	gets snapping up, caught things, stuff
HUMAN BODY	*Confronting_problem* *Taking_sides* *Part_orientational*	face, confront stance, embrace at the heart of	*Social_connection* *Body_parts* *First_rank*	contacts, close eye for, at the wrist core

Table 5: Examples of salient semantic frames (with SOURCE DOMAINS and Metaphors) in the climate change corpus (left) and the generic corpus (right).

For example, the source domain STRUGGLE (which compares mental hardships to physical ones such as fighting) includes metaphorical use of *argue* (*The ministry argued that...*) which are used in the climate corpus with similar frequency as in general texts, as well as climate-specific STRUGGLE metaphors such as *fight* (*more must be done to fight climate change*).

When comparing a topic-focused corpus against a randomized, generic corpus, log-likelihood ratio estimation allows to achieve two things: first, metaphors that are common figures of speech, used equally frequently irrespective of the text's topic, which will have a statistically insignificant differences. Second, it points towards prominent *framings* of the topic (as repeatedly used metaphors), as well as absent (but possible) framings (which metaphors are more rare than in general usage).

Table 5 shows examples of semantic frames for top source domains associated with climate change which are salient in climate discourse (left) or are avoided (right). In some source domains, this difference is particularly revealing – for example, climate metaphors in the OBJECT HANDLING source domain tend to use semantic frames denoting intentional actions such as *Attempt_suasion* (push for) or rather than more "passive" semantic frames such as *Intercepting* (caught). Interestingly, while the metaphors on both sides of the WAR source domain express fierce activity, climate change actions are not framed as a revolution (*Change_of_leadership*) or intervention (*Invading*).

In sum, semantic frames as mental templates allow to pick out specific parts of a source domain that correspond to the ways we relate to and act upon the climate crisis. However, they cannot represent the metaphorical frame on their own – as Table 5 shows, the same semantic frame (for example *Taking_sides*) can be used across different source domains, imbued with different physical imagery. Thus, discovering prevalent metaphorical frames requires both components.

4.2 Conservative vs liberal framing in immigration discourse on Twitter/X

Previous research has shown that, perhaps surprisingly, the same dehumanizing source domains are used to frame immigration by both conservatives and liberals, albeit with different frequency and intensity (Mendelsohn and Budak, 2025). Here, we study the interplay of source domains and semantic frames to shed light on the more nuanced differences of metaphor use across political camps.

4.2.1 Data and methods

To examine difference in metaphorical framing between users with different political affiliation, we use a corpus of 400K US tweets on immigration with automatically predicted metaphoricity scores across seven dehumanizing source domains and predicted scores of political affiliation (liberal vs conservative; Mendelsohn and Budak (2025). We select two well-studied domains (WATER and ANIMAL) for analysis.

We extract candidate tweets containing metaphors from our chosen source domains as explained in Section A.6, and use ConceptFrameMet to annotate the resulting 14K tweets for metaphors, their semantic frames and source domains. For precision, we only use tweets which have the same source domain annotation at both tweet (original predictions) and metaphor level (our model). Finally, since a metaphor with a WATER or ANIMAL source can refer to another, irrelevant

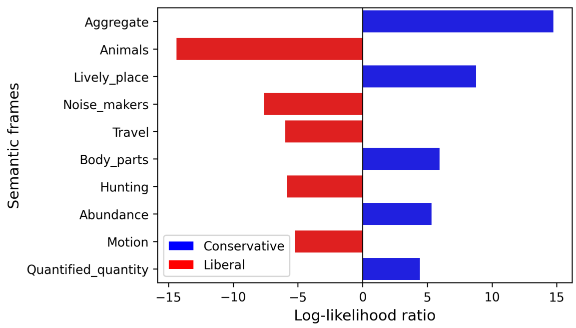

Figure 4: The five semantic frames with highest saliency in the liberal (red) and conservative (blue) data, in the ANIMAL domain.

target rather than immigration (for example, as in *left-**winged** politicians*), we further filter out tweets where the metaphor does not refer to immigrants (see Section A.7). This results in 544 tweets with source domain ANIMAL and 1,400 for WATER.

We apply log-likelihood ratio to semantic frames within the ANIMAL and WATER source domains to highlight those which are more representative of liberal vs conservative discourse. To do so, we split each subset of tweets with metaphors of a particular source domain into two parts according to their predicted ideology, and use them as C_1 and C_2, respectively.

4.2.2 Results

While many semantic frames are used by both political leanings (e.g. the top most used semantic frame in the ANIMAL source domain on is *Animals* for both sides, which is a reference to an animal or the word "animal"), their *saliency* differs. Figure 4 shows the top five most salient ANIMAL semantic frames for both leanings. The semantic frame *Animals* is salient in liberal tweets, where it is used to criticize dehumanization ("Families shouldn't be ripped apart and treated like **animals**.") Other prevalent ANIMAL metaphors in liberal tweets are based on semantic frames of *Hunting* ("ICE agents are **predators**, undocumented immigrants are **prey**"), or *Moral_evaluation* ("How is that not **predatory**?"). Conversely, the ANIMAL metaphors in conservative tweets are predominately using semantic frames like *Aggregate* ("Anytime you disrupt a rats **nest** [...] they scatter; "the **swarm** of Muslim migrants"), *Body parts* ("ILLEGALS taking [...] every benefit they can lay their **paws** on"), or *Abundance* ("the illegal immigrants will run **rampant**").

Overall, we observe that the semantic frames used by conservatives pick out more specific, "colorful", but also more negative parts of the ANIMAL source domain, while liberals often use semantic frames that show lack of agency and power, with immigrants portrayed as victims or passive (i.e. immigrants are *caged*, *hunted*, *treated like animals*).

Similarly, within the WATER source domain, the conservative features salient semantic frames like *Filling* ("**Flooding** America with illegals"; "those cities won't take **overflow** Illegal Immigrants") and *Abounding with* ("It's now a **cesspool** of crime", "CA is **saturated** with illegals"), which depict immigration as "too much" that needs to be controlled. On the other hand, the semantic frames on the liberal side have more neutral associations: *Quantified mass* ("The **waves** of desperate migrants", "the **trickle** of immigrants"), *Natural feature* ("Many different immigrant **streams**", "Migration is like the **tide**. You can put a wall up, but it will go around it."), or, again, frame immigrants as powerless victims through such semantic frames as **Killing** and **Death** ("the federales tried to **drown** families") or *Removing* ("this [policy] is in part to **flush out** immigrant parents").

5 Conclusions

Inspired by a constructionist view of metaphor theory, which posits that metaphors are bounded by their linguistic expression, we proposed a framework, ConceptFrameMet, that models metaphors through both their source domains and semantic frames and show how salient semantic frames within source domains allow to capture nuanced differences in framing evoked by the metaphors. We release a pre-trained model with predicts metaphors, their source domains and semantic frames, as well as a statistical model for estimating the saliency of source domains and semantic frames within a particular topic or discourse. We demonstrate the potential of our framework by applying it to two different tasks – *discovering* source domains and semantic frames used in framing climate change news, and *highlighting nuanced differences* in semantic frames used by liberals and conservatives within the same metaphorical source domain. Future work can include integrating it with framework such as MetaNet (Dodge et al., 2015) to enable more sophisticated analysis of semantic frames that support complex, extended discourse metaphors.

5.1 Limitations

Our study has several limitations. First, we use a model trained on a set of source domains used in the LCC corpus (Mohler et al., 2016), which might not be comprehensive enough to cover new topics that are very different from the ones occurring in it. Thus, a larger – or even open – set of source domains is needed, and a method which is sensitive enough to differentiate between a large number of source domains. Relatedly, the question of distinguishing between related or semantically similar source domains (Shutova and Teufel, 2010) is still unanswered, though in Section 3.1.2 we show an example of how semantic frames can serve as a tool to make such distinctions in a more principled way. Finally, in this study we only look at individual semantic frames within a particular domain, and do not consider how they interplay and link with each other in the same document to create more powerful extended metaphors. Thus, we hope to connect our FrameNet-frames based analysis – which operates at lexical level – with the MetaNet (Dodge et al., 2015) framework, which encodes the relationships between the components of the source domain.

References

Kaikai An, Ce Zheng, Bofei Gao, Haozhe Zhao, and Baobao Chang. 2023. Coarse-to-fine dual encoders are better frame identification learners. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 13455–13466, Singapore. Association for Computational Linguistics.

Daniel Baleato Rodríguez, Verna Dankers, Preslav Nakov, and Ekaterina Shutova. 2023. Paper bullets: Modeling propaganda with the help of metaphor. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 472–489, Dubrovnik, Croatia. Association for Computational Linguistics.

Julia Bingler, Mathias Kraus, Markus Leippold, and Nicolas Webersinke. 2023. How cheap talk in climate disclosures relates to climate initiatives, corporate emissions, and reputation risk. Working paper, Available at SSRN 3998435.

Julia Birke and Anoop Sarkar. 2006. A clustering approach for nearly unsupervised recognition of nonliteral language. In *11th Conference of the European Chapter of the Association for Computational Linguistics*, pages 329–336, Trento, Italy. Association for Computational Linguistics.

Amber Boeynaems, Christian Burgers, Elly A Konijn, and Gerard J Steen. 2017. The effects of metaphorical framing on political persuasion: A systematic literature review. *Metaphor and Symbol*, 32(2):118–134.

Joanne Boisson, Arif Mehmood, and Jose Camacho-Collados. 2025. METAPHORSHARE: A dynamic collaborative repository of open metaphor datasets. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (System Demonstrations)*, pages 509–521, Albuquerque, New Mexico. Association for Computational Linguistics.

Britta C Brugman, Christian Burgers, and Barbara Vis. 2019. Metaphorical framing in political discourse through words vs. concepts: A meta-analysis. *Language and Cognition*, 11(1):41–65.

Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan, and Nanyun Peng. 2021. MERMAID: Metaphor generation with symbolism and discriminative decoding. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4250–4261, Online. Association for Computational Linguistics.

Jieyu Chen, Kathleen Ahrens, and Chu-Ren Huang. 2022. Framing legitimacy in CSR: A corpus of Chinese and American petroleum company CSR reports and preliminary analysis. In *Proceedings of the First Computing Social Responsibility Workshop within the 13th Language Resources and Evaluation Conference*, pages 24–34, Marseille, France. European Language Resources Association.

Minjin Choi, Sunkyung Lee, Eunseong Choi, Heesoo Park, Junhyuk Lee, Dongwon Lee, and Jongwuk Lee. 2021. Melbert: Metaphor detection via contextualized late interaction using metaphorical identification theories. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1763–1773.

Jacob Devasier, Yogesh Gurjar, and Chengkai Li. 2024. Robust frame-semantic models with lexical unit trees and negative samples. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6930–6941, Bangkok, Thailand. Association for Computational Linguistics.

Ellen Dodge, Jisup Hong, and Elise Stickles. 2015. MetaNet: Deep semantic automatic metaphor analysis. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 40–49, Denver, Colorado. Association for Computational Linguistics.

Elisabeth El Refaie. 2001. Metaphors we discriminate by: Naturalized themes in austrian newspaper articles about asylum seekers. *Journal of Sociolinguistics*, 5(3):352–371.

Robert M Entman. 1993. Framing: Towards clarification of a fractured paradigm. *McQuail's reader in mass communication theory*, 390:397.

Ethan Fast and Eric Horvitz. 2017. Long-term trends in the public perception of artificial intelligence. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

Charles J Fillmore and Collin F Baker. 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, volume 6, pages 59–64.

Bálint Forgács and Csaba Pléh. 2022. The fluffy metaphors of climate science. In *Metaphors and analogies in sciences and humanities: Words and worlds*, pages 447–477. Springer.

Maucha Gamonal. 2022. A descriptive study of metaphors and frames in the multilingual shared annotation task. In *Proceedings of the Workshop on Dimensions of Meaning: Distributional and Curated Semantics (DistCurate 2022)*, pages 1–7, Seattle, Washington. Association for Computational Linguistics.

Mengshi Ge, Rui Mao, and Erik Cambria. 2022. Explainable metaphor identification inspired by conceptual metaphor theory. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 10681–10689.

Mengshi Ge, Rui Mao, and Erik Cambria. 2023. A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, 56(Suppl 2):1829–1895.

Jonathan Gordon, Jerry Hobbs, Jonathan May, Michael Mohler, Fabrizio Morbini, Bryan Rink, Marc Tomlinson, and Suzanne Wertheim. 2015. A corpus of rich metaphor annotation. In *Proceedings of the Third Workshop on Metaphor in NLP*, pages 56–66, Denver, Colorado. Association for Computational Linguistics.

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

Yolanda Guan and Winnie Huiheng Zeng. 2024. Changes in the sentiments and metaphors in COVID-19 news discourse (2019-2024). In *Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation*, pages 810–819, Tokyo, Japan. Tokyo University of Foreign Studies.

Carl G Herndl and Stuart C Brown. 1996. *Green culture: Environmental rhetoric in contemporary America*.

Jingyi Huang and Ming Liu. 2025. Metaphorical framing of climate change in chinese and american news media: A corpus-assisted discourse study. *Critical Arts*, pages 1–20.

Hyeju Jang, Keith Maki, Eduard Hovy, and Carolyn Rosé. 2017. Finding structure in figurative language: Metaphor detection with topic-based frames. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 320–330, Saarbrücken, Germany. Association for Computational Linguistics.

Mark Jones. 1992. Generating a specific class of metaphors. In *30th Annual Meeting of the Association for Computational Linguistics*, pages 321–323.

Rohan Joseph, Timothy Liu, Aik Beng Ng, Simon See, and Sunny Rai. 2023. NewsMet : A 'do it all' dataset of contemporary metaphors in news headlines. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10090–10104, Toronto, Canada. Association for Computational Linguistics.

Oleksandr Kapranov. 2017. Conceptual metaphors associated with climate change in corporate reports in the fossil fuels market: Two perspectives from the united states and australia. In *The Role of Language in the Climate Change Debate*, pages 90–109. Routledge.

George Lakoff and Mark Johnson. 2008. *Metaphors we live by*. University of Chicago press.

Mark J Landau, Daniel Sullivan, and Jeff Greenberg. 2009. Evidence that self-relevant motives and metaphoric framing interact to influence political and social attitudes. *Psychological science*, 20(11):1421–1427.

Chee Wee (Ben) Leong, Beata Beigman Klebanov, and Ekaterina Shutova. 2018. A report on the 2018 VUA metaphor detection shared task. In *Proceedings of the Workshop on Figurative Language Processing*, pages 56–66, New Orleans, Louisiana. Association for Computational Linguistics.

Yanlin Li, Jing Chen, Kathleen Ahrens, and Chu-Ren Huang. 2024. The evolving use of WAR metaphors in businesswomen-focused media discourse. In *Proceedings of the 38th Pacific Asia Conference on Language, Information and Computation*, pages 1377–1386, Tokyo, Japan. Tokyo University of Foreign Studies.

Yucheng Li, Shun Wang, Chenghua Lin, Frank Guerin, and Loic Barrault. 2023. FrameBERT: Conceptual metaphor detection with frame embedding learning. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1558–1563, Dubrovnik, Croatia. Association for Computational Linguistics.

Emmy Liu, Chenxuan Cui, Kenneth Zheng, and Graham Neubig. 2022. Testing the ability of language models to interpret figurative language. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4437–4452, Seattle, United States. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Rui Mao, Xiao Li, Kai He, Mengshi Ge, and Erik Cambria. 2023. MetaPro online: A computational metaphor processing online system. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 127–135, Toronto, Canada. Association for Computational Linguistics.

Julia Mendelsohn. 2024. *Theory-grounded Computational Analysis of Political Framing in Online Media*. Phd dissertation, University of Michigan.

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, Yulia Tsvetkov, and Dan Jurafsky. 2020. A framework for the computational linguistic analysis of dehumanization. *Frontiers in artificial intelligence*, 3:55.

Haohan Meng, Xiaoyu Li, and Jinhua Sun. 2025. Large language models prompt engineering as a method for embodied cognitive linguistic representation: a case study of political metaphors in trump's discourse. *Frontiers in Psychology*, 16:1591408.

Saif Mohammad, Ekaterina Shutova, and Peter Turney. 2016. Metaphor as a medium for emotion: An empirical study. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, pages 23–33, Berlin, Germany. 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.

Brigitte Nerlich and Rusi Jaspal. 2012. Metaphors we die by? geoengineering, metaphors, and the argument from catastrophe. *Metaphor and symbol*, 27(2):131–147.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15407–15428, Bangkok, Thailand. Association for Computational Linguistics.

Vinodkumar Prabhakaran, Marek Rei, and Ekaterina Shutova. 2021. How metaphors impact political discourse: A large-scale topic-agnostic study using neural metaphor detection. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 503–512.

Sunny Rai and Shampa Chakraverty. 2020. A survey on computational metaphor processing. *ACM Computing Surveys (CSUR)*, 53(2):1–37.

Paul Rayson and Roger Garside. 2000. Comparing corpora using frequency profiling. In *The Workshop on Comparing Corpora*, pages 1–6, Hong Kong, China. Association for Computational Linguistics.

Sebastian Reimann and Tatjana Scheffler. 2025. Using large language models to perform MIPVU-inspired automatic metaphor detection. In *Proceedings of the 2nd Workshop on Analogical Abstraction in Cognition, Perception, and Language (Analogy-Angle II)*, pages 10–21, Vienna, Austria. Association for Computational Linguistics.

Zachary Rosen. 2018. Computationally constructed concepts: A machine learning approach to metaphor interpretation using usage-based construction grammatical cues. In *Proceedings of the Workshop on Figurative Language Processing*, pages 102–109, New Orleans, Louisiana. Association for Computational Linguistics.

Josef Ruppenhofer, Michael Ellsworth, Miriam RL Petruck, and Christopher R Johnson. Framenet ii: Extended theory and practice.

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.

Bertram T Scheufele and Dietram A Scheufele. 2010. Of spreading activation, applicability, and schemas: Conceptual distinctions and their operational implications for measuring frames and framing effects. In *Doing News Framing Analysis*, pages 126–150. Routledge.

Elena Semino, Zsófia Demjén, and Jane Demmen. 2018. An integrated approach to metaphor and framing in cognition, discourse, and practice, with an application to metaphors for cancer. *Applied linguistics*, 39(5):625–645.

Meghdut Sengupta, Milad Alshomary, and Henning Wachsmuth. 2022. Back to the roots: Predicting the source domain of metaphors using contrastive learning. In *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*, pages 137–142, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Meghdut Sengupta, Roxanne El Baff, Milad Alshomary, and Henning Wachsmuth. 2024. Analyzing the use of metaphors in news editorials for political framing. 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 3621–3631, Mexico City, Mexico. Association for Computational Linguistics.

Ekaterina Shutova and Simone Teufel. 2010. Metaphor corpus annotated for source-target domain mappings. In *LREC*, volume 2, pages 2–2.

Jorunn Skinnemoen. 2009. Metaphors in climate change discourse. Master's thesis.

Gerard J Steen, Aletta G Dorst, Tina Krennmayr, Anna A Kaal, and J Berenike Herrmann. 2010. A method for linguistic metaphor identification.

Kevin Stowe, Nils Beck, and Iryna Gurevych. 2021a. Exploring metaphoric paraphrase generation. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 323–336, Online. Association for Computational Linguistics.

Kevin Stowe, Tuhin Chakrabarty, Nanyun Peng, Smaranda Muresan, and Iryna Gurevych. 2021b. Metaphor generation with conceptual mappings. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6724–6736, Online. Association for Computational Linguistics.

Karen Sullivan. 2013. *Frames and constructions in metaphoric language*. John Benjamins Publishing Company.

Karen Sullivan. 2023. Three levels of framing. *Wiley Interdisciplinary Reviews: Cognitive Science*, 14(5):e1651.

Karen Sullivan. 2025. *Metaphors and Constructions*, page 129–146. Cambridge Handbooks in Language and Linguistics. Cambridge University Press.

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A Smith. 2017. Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. *arXiv preprint arXiv:1706.09528*.

Yuan Tian, Minzheng Wang, Nan Xu, and Wenji Mao. 2025. ImaRA: An imaginative frame augmented method for low-resource multimodal metaphor detection and explanation. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 3953–3967, Albuquerque, New Mexico. Association for Computational Linguistics.

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.

Oseremen Uduehi and Razvan Bunescu. 2024. An expectation-realization model for metaphor detection. In *Proceedings of the 4th Workshop on Figurative Language Processing (FigLang 2024)*, pages 79–84, Mexico City, Mexico (Hybrid). Association for Computational Linguistics.

Tony Veale. 2016. Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 34–41, San Diego, California. Association for Computational Linguistics.

Dian Wang, Yang Li, Suge Wang, Xin Chen, Jian Liao, Deyu Li, and Xiaoli Li. 2025. Ckemi: Concept knowledge enhanced metaphor identification framework. *Information Processing & Management*, 62(1):103946.

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.

Nicolas Webersinke, Mathias Kraus, Julia Anna Bingler, and Markus Leippold. 2021. ClimateBERT: A pretrained language model for climate-related text. *arXiv preprint arXiv:2110.12010*.

Kaiwen Yang and Ya Sun. 2025. A tale of identities: Environmental identities based on a deliberate metaphor analysis of us energy companies' social media. *Journal of Business and Technical Communication*, 39(2):149–190.

Fengying Ye, Shanshan Wang, Lidia S Chao, and Derek F Wong. 2025. Unveiling llms' metaphorical understanding: Exploring conceptual irrelevance, context leveraging and syntactic influence. *arXiv preprint arXiv:2510.04120*.

Shenglong Zhang and Ying Liu. 2022. Metaphor detection via linguistics enhanced Siamese network. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4149–4159, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jörg Zinken, Iina Hellsten, and Brigitte Nerlich. 2008. Discourse metaphors. *Body, language and mind*, 2:363–385.

A Appendix

A.1 Pre-training models finetuning details

A.1.1 Semantic frame prediction

Fine-tuning. We use input length of 256, batch size of 32, Adam W optimization, and 2 warm-up epochs. All models are trained for the maximum of 10 epochs with patience of 2 as determined by absence of improvement in terms of binary-F1. The starting learning rate is $2e^{-5}$ with linear learning rate scheduler and weight decay of 0.1.

A.1.2 Source domain prediction

Data. We use the large version (ENGLISH large) of the LCC Metaphor Dataset (large version) (Mohler et al., 2016). To ensure high quality of training data, we only use human-validated subsets of the dataset (ANN, REC, SYS), and only the samples where at least half of the annotators regarding the source domain, or there is only one annotation. We also remove those source domains which have less than 3 samples, to ensure valid evaluation and testing. This results in 16722 samples, out of which 11154 have single annotations, and 5568 are annotated by at least two people (with a high agreement in terms of Krippendorf's α 0.76 and Fleiss κ of 0.76).

Fine-tuning details. All models have the same hyperparameters in terms of input length (256), batch size of 32, Adam W optimization, and 2 warm-up epochs. All models are trained for the maximum of 30 epochs with patience of 5 as determined by absence of improvement in terms of macro-F1. The starting learning rate is $2e^{-5}$ with linear learning rate scheduler and weight decay of 0.1.

A.1.3 Metaphor prediction

For fair comparison with Melbert, we reproduce fine-tuning parameters used in the original implementation (Choi et al., 2021). In particular, we use the learning rate of 3e-05 with 2 warm-up epochs and the drop ratio of 0.2, and train for 3 epochs.

A.2 Classifier prompts

A.2.1 Semantic frame classification

We use the following prompt, where **labels** is substituted with the list of semantic frames and their definitions:

```
You are a linguist specializing in
semantic frames (FrameNet).

Please choose the semantic frame out of
the following list: labels. Do not use
any other labels, and do not change the
wording of the label. Do not remove "_"
in the label if it exists.

Please return a json object which consists
of the following field:

"frame": one of the values from the list.

Do not output anything else. Do not output
any reasoning steps or explanations.
```

A.2.2 Source domain classification

We use the following prompt, which is based on human annotation instructions (see Section A.8), where **labels** is substituted with the list of source domains:

```
You are a linguist specializing in
metaphors. You will be given a metaphor
and a sentence it occurs in. You will be
asked to identify the source domain of
the metaphor.

A metaphor is when you describe something
by saying it is something else, even
though it is not literally true. For
example, we can say "They are forced to
make a decision" while there is no actual
physical force applied to them.

When using a metaphor, we are carrying
over associations from a more tangible and
specific source domain (such as physical
force or pressure) to a more abstract
domain (such as obligation).

In this study, you will need to identify
the source domain of the metaphor. It is
helpful to remember that the source domain
is usually a more specific, physical
thing (force, pressure) while the target
domain is more abstract (obligation).

Please choose the source domain out of
the following list: labels. Do not use
any other labels, and do not change the
wording of the label, Do not remove "_"
in the label if it exists.

Please return a json object which consists
of the following field:

"source": one of the values from the list.

Do not output anything else. Do not output
any reasoning steps or explanations.
```

A.2.3 Metaphor prediction

We use the MIP protocol for the prompt, which our preliminary experiments have shown to be more effective than other metaphor identification protocols such as SPV or CMT.

```
You are a linguist specializing in
metaphors. You will be given a sentence
and asked to find all metaphors in it.
Please extract all text spans that have a
```

```
metaphorical meaning.

Go through the following steps to
determine if a word is used in a
metaphorical meaning:

1. Read the entire sentence to establish
a general understanding of the meaning.
2. (a) Establish the word's meaning in
context, that is, how it applies to an
entity, relation, or attribute in the
situation evoked by the text (contextual
meaning). Take into account what comes
before and after the lexical unit.
(b) Determine if the target word has
a more basic contemporary meaning in
other contexts than the one in the given
context. For our purposes, basic meanings
tend to be
–More concrete; what they evoke is easier
to imagine, see, hear, feel, smell, and
taste.
–Related to bodily action.
–More precise (as opposed to vague)
–Historically older.
Basic meanings are not necessarily the
most frequent meanings of the target
word.
(c) If the target word has a more
basic current–contemporary meaning in
other contexts than the given context,
decide whether the contextual meaning
contrasts with the basic meaning but can
be understood in comparison with it.
3. If yes, the word is metaphorical.
Otherwise it is literal.

Please return a json object which consists
of the following field:

metaphors: a list of extracted metaphor
spans.

Do not output any explanations or
reasoning steps.
```

To evaluate the extracted spans, we check if they overlap with golden spans in the dataset. We consider partial overlaps to be correct.

A.3 Source prediction performance depending on the class frequency

Figure 5 compared performance of SEP-style RoBERTa models with and without semantic frames on test set for classes with different frequency in the training set.

A.4 Distribution of mentions of "climate change" and "global warming" across 1986-2020

Figure 6 shows distribution of mentions of "climate change" and "global warming" in New York Times over the period from 1986 to 2020.

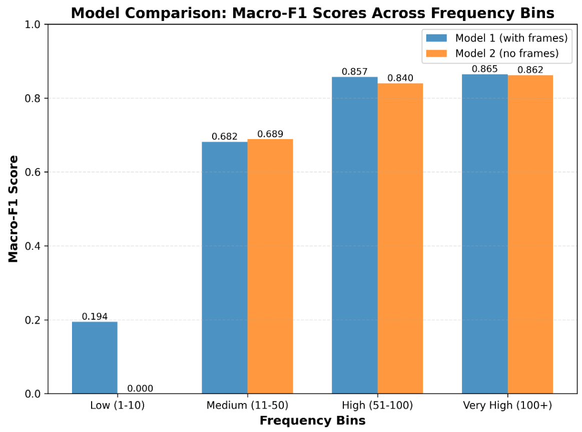

Figure 5: Performance comparison on test set for classes with different frequency in the training set.

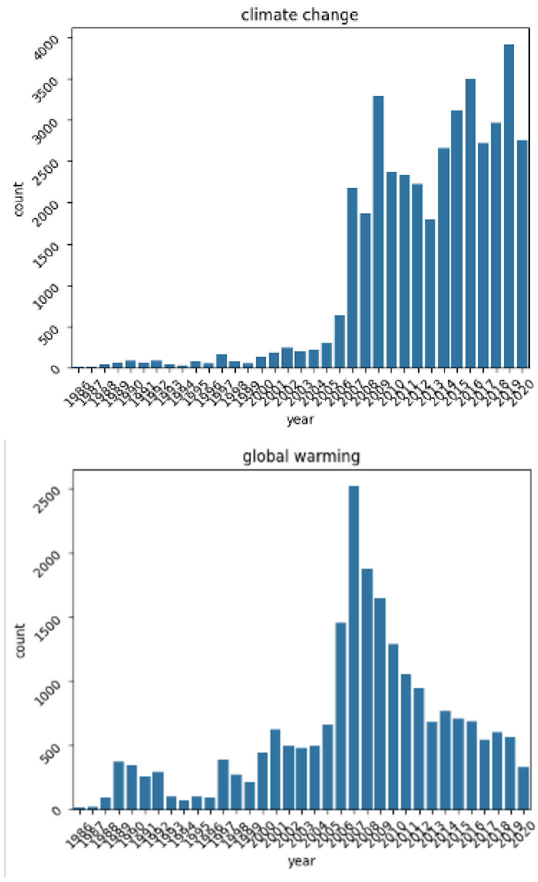

Figure 6: Distribution of mentions of "climate change" and "global warming" across 1986-2020

A.5 Top source domains in climate corpus

Table 6 shows the most salient source domains in NYT climate corpus, together with their examples, alternative names, and selected theoretical studies that identified them.

A.6 Immigration source domains

The corpus provided by (Mendelsohn and Budak, 2025) contains tweets automatically annotated with predicted metaphoricity scores across seven dehumanization source domains. The authors also release a smaller dataset with 200 tweets per source domain, where each tweet is judged by 10 human annotators as containing or not containing a metaphor from that particular domain.

As the larger dataset only contains automatically predicted probabilities, we convert them to (potential) source domain labels as follows. We regard a tweet t as potentially containing a metaphor with a particular source domain s if the predicted probability of that source domain $P(s|t)$ is larger than the average metaphoricity score $\bar{m}_s = \frac{1}{|T_s|}\sum_{t' \in T_s} m_{t'}$ of tweets that were judged by annotators as belonging to that source domain (these cut-off scores are in Section A.6).

Table 7 below shows the source domains used in metaphorical framing of tweets about immigration, their definitions, the average metaphoricity scores from the annotated corpus by (Mendelsohn and Budak, 2025), the number of tweets from the unlabeled part of that corpus with metaphoricity scores about that cut off, as well as the final number of tweets used for analysis after filtering

A.7 Filtering out metaphors which refer to non-relevant target domains

To filter out metaphors which have the source domain we are interested in but the wrong (irrelevant) target, we first use Claude Sonnet 4.5 with the prompt below, where we ask it to resolve the target which the specific metaphor refers to:

```
You are a linguist specializing in
metaphors. You will be given a sentence
and a metaphor in it. Identify and out
out the target of the metaphor, i.e. the
entity, person, or object it refers to.
For example, given the following input:

"text": "Immigrants are flooding
into our country and ruining our economy"
"metaphor": "flooding"

output the following:

"target": "immigrants"

Make sure to use json format with
this field. Do not output anything else,
no explanations! You are NOT allowed to
say "no metaphor is present" since the
metaphor is given to you!
```

In this way, we collect a list of target domains used in tweets, such as "undocumented immigrants", "illegals", "Liberals", "economy", or "Trump", and filter it out manually to collect a list of expressions that are likely to refer to the target domain we are interested in (immigration). In particular, we arrive at the following list of substrings which then match against the predicted target domains: 'immi', 'illegal', 'alien', 'ICE', 'migr', 'foreign', 'refug', 'detention', 'border', 'asylum', 'famil'.

A.8 Annotation Task Details

A.8.1 Task Overview

We collected human annotations for metaphor source domains using a custom web-based annotation platform.[5] Annotators were presented with sentences containing highlighted metaphors and asked to identify up to 3 source domains from a provided list of options for each metaphor. The annotation interface included a mandatory 10 second reading period before options appeared, and copy-paste was not allowed, designed to encourage careful consideration of the metaphorical usage in context and to avoid AI-powered tools.

A.8.2 Annotator Recruitment

Annotators were recruited through Prolific (https://www.prolific.com/), a crowdsourcing platform commonly used for academic research. The study collected participants' Prolific IDs, study IDs, and session IDs to enable proper tracking and compensation. All participants were required to provide informed consent through their Prolific ID before proceeding to the annotation task. Participants were required to be native English speakers, with English listed as their first language. Eligible annotators reported their country of birth as either the United Kingdom or the United States. To ensure a high level of reading comprehension and annotation quality, participants were additionally required to have completed at least an undergraduate degree (BA/BSc or equivalent), with many

[5] Available at https://metaphor-annotation-source.onrender.com/

Source domain	Alternative names	Examples	Reference
STRUGGLE	CHALLENGE	fight, confront	Kapranov (2017)
WAR	BATTLE	battle, war, on the front lines	Skinnemoen (2009)
OBJECT HANDLING	OBJECT	push (for), press, handle, take (action)	Herndl and Brown (1996)
HUMAN BODY	BODY	face, confront, embrace, at the heart of	Nerlich and Jaspal (2012)
PATHWAY	PATH	passing, path	Yang and Sun (2025)
BUILDING	CONSTRUCTION	build, lay a foundation	Huang and Liu (2025)
FURNISHINGS		climategate, take a back seat	Forgács and Pléh (2022)
VISION		view, presage	
FORWARD MOVEMENT		ahead, progress, gain	Skinnemoen (2009)
BACKWARD MOVEMENT		backwards, rollback	Skinnemoen (2009)

Table 6: Top 10 salience source domains in NYT climate corpus, together with their alternative names in the literature, examples and references. For brevity we provide only one reference per source domain.

Source domain	Examples	Avg metaphoricity score	# of candidates	# after filtering
Animal	shelter, cage, swarm	0.3045	9845	544
Water	flood, tide, pour	0.4375	3661	1402

Table 7: Statistics for ANIMAL and WATER source domains

holding graduate or doctoral degrees. Only Prolific users with an approval rate between 98% and 100% were permitted to participate.

A.8.3 Annotation Instructions

Figure 7 features the instructions provided to annotators during the interactive tutorial phase.

A.8.4 Quality Assurance

We implemented a multi-stage quality assurance protocol to ensure high annotation reliability. Prior to beginning the main annotation task, participants were shown five example annotations and then completed five comprehension (attention check) questions with known correct answers. To pass, participants were required to answer at least three out of five questions correctly (i.e., a maximum of two errors). Participants who failed the comprehension check were permitted one retry after reviewing the task instructions. A second failure resulted in permanent rejection and participants were instructed to return their submission on Prolific.

During the annotation task, participants were required to complete each sample sequentially before proceeding to the next. For each sample, annotators could select a maximum of three source domains. They could also select *No Metaphor*, which was mutually exclusive with all other options, or *Other Domain*, which prompted them to manually specify a domain not listed among the predefined options.

The annotation platform explicitly tracked three rejection scenarios:

1. **First comprehension check failure**: participants were allowed to retry after reviewing the instructions.
2. **Second comprehension check failure**: participants were permanently rejected and instructed to return their submission on Prolific.
3. **Incomplete submissions**: participants who did not complete all 50 annotations in their assigned batch were automatically instructed to return their submission.

Participants were informed during the consent process that failing comprehension checks would result in submission return or rejection, and that repeated failures would lead to ineligibility for future tasks within the project unless manually re-approved.

A.8.5 Compensation

Each annotation batch was completed by four independent annotators. Participants were allotted up to 45 minutes to complete a batch of 50 annotations. Compensation was set at an hourly rate of £12, in line with Prolific's recommended fair pay guidelines and well above the local standard minimum wage.

A.8.6 Annotation agreement per batch

Table 8 shows averaged pair-wise annotation agreement within each batch, as well as overall (mean) agreement across batches

Batch	Agreement rate
1	0.65
2	0.63
3	0.71
4	0.73
5	0.76
6	0.53
7	0.84
8	0.66
Mean	0.69

Table 8: Per batch IAA rate

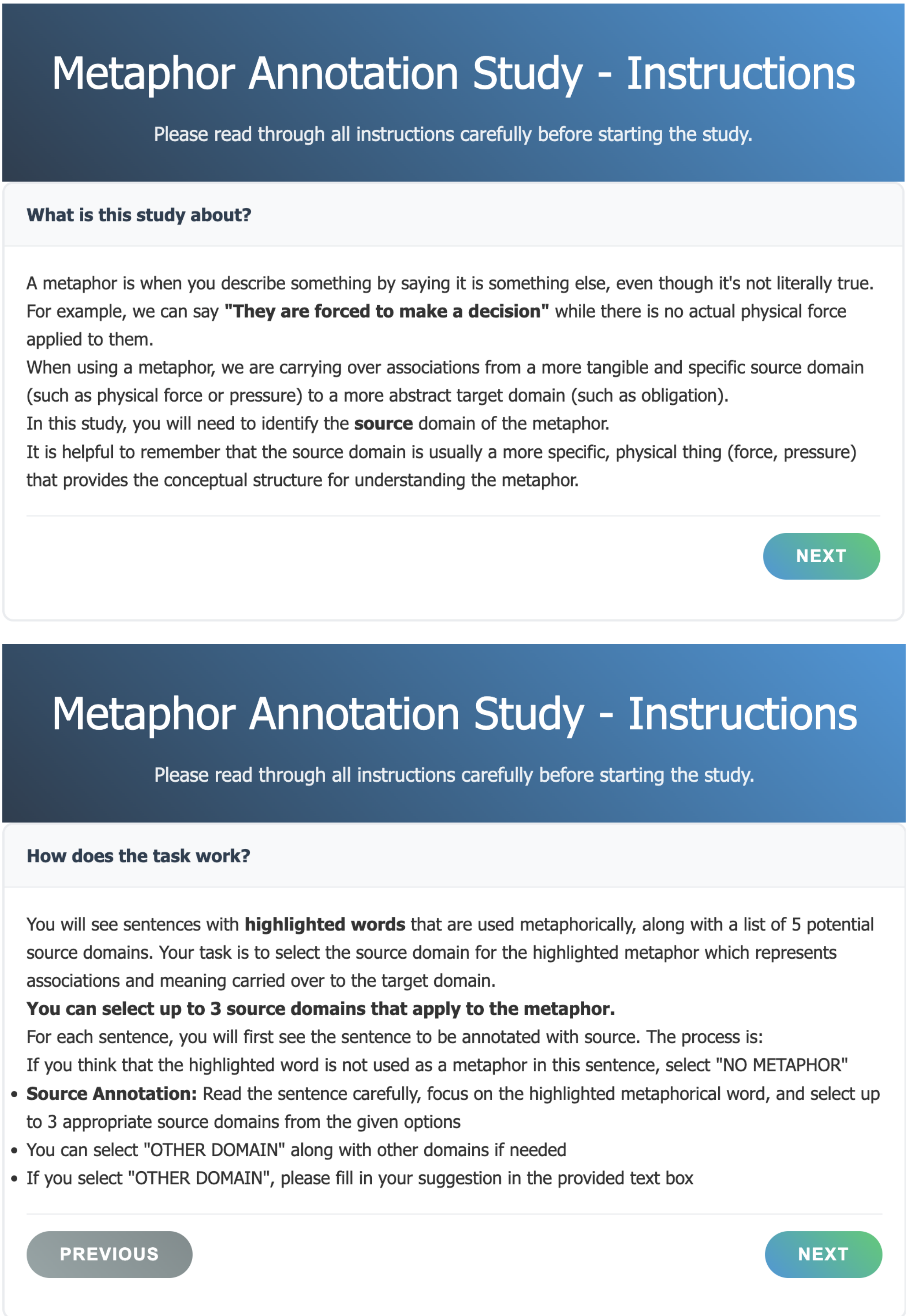

Figure 7: Screenshots of the annotation instruction interface. Annotators were guided through an interactive tutorial explaining (top) the concept of metaphor source domains and (bottom) the annotation task workflow before beginning comprehension checks.