Analyzing the Use of Metaphors in News Editorials for Political Framing

Meghdut Sengupta[1] **Roxanne El Baff** [2]
Milad Alshomary[3] * **Henning Wachsmuth**[1]
Leibniz Universität Hannover, Germany[1] German Aerospace Center (DLR), Germany[2]
Columbia University, USA[3]
{m.sengupta, h.wachsmuth}@ai.uni-hannover.de[1]
roxanne.elbaff@dlr.de[2], ma4608@columbia.edu[3]

Abstract

Metaphorical language is a pivotal element in the realm of political framing. Existing work from linguistics and the social sciences provides compelling evidence regarding the distinctiveness of conceptual framing for political ideology perspectives. However, the nature and utilization of metaphors and the effect on audiences of different political ideologies within political discourses are hardly explored. To enable research in this direction, in this work we create a dataset, originally based on news editorials and labeled with their persuasive effects on liberals and conservatives and extend it with annotations pertaining to metaphorical usage of language. To that end, first, we identify all single metaphors and composite metaphors. Secondly, we provide annotations of the source and target domains for each metaphor. As a result, our corpus consists of 300 news editorials annotated with spans of texts containing metaphors and the corresponding domains of which these metaphors draw from. Our analysis shows that liberal readers are affected by metaphors, whereas conservatives are resistant to them. Both ideologies are affected differently based on the metaphor source and target category. For example, liberals are affected by metaphors in the *Darkness & Light* (e.g., *death*) source domains, where as the source domain of *Nature* affects conservatives more significantly.

1 Introduction

Metaphorical language is dominant in various domains of everyday life (Lakoff and Johnson, 2020). Particularly, politicians widely frame their discourses by means of employing metaphors (Lakoff, 1995; Mio, 1997; Chilton and Ilyin, 1993; Charteris-Black, 2009).

Inspired from the aforementioned research regarding metaphorical usages in political domains,

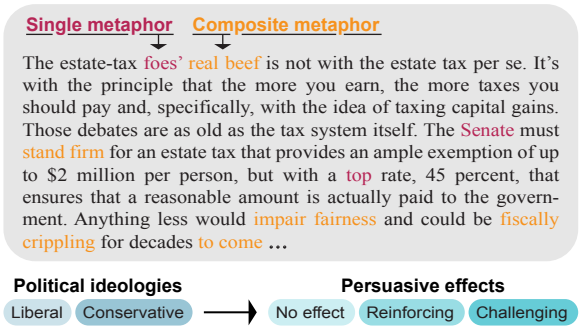

Figure 1: Text snippet from of our data from the corpus: Single and composite metaphors are marked in red and blue respectively. For each editorial we have two political categories, namely liberal and conservative. Each of the two categories can have either *No effect*, *challenging*, or *reinforcing* effect on the reader

studies related to modelling metaphors with respect to political framing has been investigated like Cabot et al. (2020). In their work, they modelled a multitask learning setup to predict political perspective of news articles (Baly et al., 2020), party affiliations of politicians (Biessmann et al., 2016), and framing dimensions of socioeconomic issues such as *gun-control* (Card et al., 2015) as the main task, while predicting metaphors (Gong et al., 2020) and emotions (Troiano et al., 2023) act as the auxiliary tasks.

To that end they employ five different datasets from the domains of political bias (Li and Goldwasser, 2019), political affiliation (Li and Goldwasser, 2019), framing (Card et al., 2015), metaphors (Steen, 2010), and emotions labelled on tweet data (Mohammad et al., 2018).

While different areas of language involve different degrees of metaphor usage, news editorials utilize metaphors extensively as a rhetorical device to better reach their audiences (Farrokhi and Nazemi, 2015). Recent work by Joseph et al. (2023), among others, draws on annotations of metaphors in content related to politics and others, but only with news headlines as data.

*work carried out as a research assistant at Leibniz Universität Hannover, Germany

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
June 16-21, 2024 ©2024 Association for Computational Linguistics

Hence, besides the need of having a dataset at the intersection of metaphors with political discourses and framing, an under-explored dimension in this context is the persuasive effect of a political discourse on the readers. El Baff et al. (2018) determines persuasive effect of an editorial, based on whether the editorial challenged the reader's stance by making them rethink it, reinforced their stance by helping them argue better, or was ineffective . In their work, for each ideology - liberal and conservative - there are three effects of persuasiveness - *reinforcing*, *challenging*, and *no effect*.

This paper explores where and how metaphors are used for political framing concerning persuasiveness, focusing on news editorials. We do this in conjunction with employing the conceptual domains in metaphorical meaning construction (Lakoff and Johnson, 2008), namely, the *source* - the concept domain from where the meaning is derived in a metaphorical sentence and *target domain* - the concept domain which is explained by metaphor in the sentence (see section 2).

To that end, we re-annotate an existing corpus of news editorials from the New York Times (El Baff et al., 2018) by two levels of annotations. Firstly, we identify and annotate the metaphors with the metaphor identification procedure by Steen (2010), and secondly, we annotate each metaphor with their source and target domains, which we reuse from the work by Shutova et al. (2010) and Gordon et al. (2015).

We do two pilot studies each with 3 samples of editorials chosen randomly from the corpus by El Baff et al. (2018) with two annotators. For the main study, we employed an additional annotator. Hence a total of 3 annotators carried out the main annotation on 300 randomly sampled news editorials from the same corpus. The average number of paragraphs of each editorial is 7.3. The total number of metaphors present in the corpus are 12006 with 8353 single metaphors (70 %) and 3653 composite metaphors (30 %). We have an overall 84 source domains and 59 target domains distributed across out dataset. Figure 1 shows a snippet of a sample annotation.

Our findings suggest that the political stance of liberals related to political ideologies are affected significantly by metaphorical language used in the news editorials than conservatives. For analysing the effect of source and target domains on the political stance by these two ideologies, we cluster the source and target domains into 14 ontological categories based on Gordon et al. (2015). Our results based on this analysis indicate that the metaphors and source domains have adverse persuasive effects on the political stance of the liberals, especially from the categories of *Nature*, and *Darkness & Light*.

In summary, our main contributions are as follows:

- We create a dataset to study metaphorical usages in political news editorials.
- We analyze the relationship of the metaphorical usage in news editorials with respect to the perceived effect (reinforcing, challenging, no-effect) by conservative and liberal readers by answering the following two questions: (1) Does the usage of metaphors affect readers with different ideologies differently? (2) Which source domains correlate to the persuasive effect of editorials perceived by each ideology?

2 Related Work

Metaphors are a linguistic tool used to convey implicit meaning, while their literal counterparts in the context of a sentence are explicit. This gives rise to a phenomenon known as *meaning shift* (Lakoff and Johnson, 2008). A metaphor, represented by a specific word or phrase in a sentence, facilitates this shift. It involves extracting meaning from an abstract conceptual domain, referred to as the *source domain*, and projecting it onto a concrete conceptual domain, referred to as the *target domain*. Consider this metaphorical sentence:

"Wages have been stagnating through much of the current economic cycle."

Here, "stagnating" is metaphorical, since *stagnation* as a concept cannot be experienced by a non-physical entity such as *wage* in the physical realms. The intended meaning is manifested by establishing a mapping of two conceptual domains, drawing the meaning from a source domain (*movement*) and projecting it into a target domain (*money*).

Foundational concepts such as source and target domains have been considered in the development of the subsequent set of datasets (Shutova et al., 2010; Chakrabarty et al., 2021; Zhang et al., 2021). Recently, Sengupta et al. (2022) used contrastive learning to predict the source domains of

metaphors, and Sengupta et al. (2023) extended this by modelling the aspects highlighted by metaphors, in a given context, in a multitask setting. Their work showed that for predicting highlighted aspects of metaphors, incorporating the information of the source domains of the concerned metaphor improves model performance and vice versa.

However, their work is a first step towards interpretation of metaphors via exploiting metaphorical components and limits itself from establishing a direct connection of the conceptual domains to a real-world setting (like political discourses). Unlike their work, our dataset provides a groundwork for establishing a direct connection of the components constructing metaphorical meaning to persuasiveness in political discourses.

The task of metaphor detection, aiming to distinguish between literal and metaphorical uses of specific words within sentences, has been studied in NLP research extensively (Shutova et al., 2010). To that end, various datasets have been utilized in the course of this research line (Steen, 2010; Gordon et al., 2015; Mohler et al., 2016; Do Dinh et al., 2018).

Our annotation guideline builds on top of the metaphor identification system initially designed to construct the VU Amsterdam Metaphor Corpus (Steen, 2010) to identify the metaphors in the dataset. To annotate the source and target domains, we reuse the same provided by Shutova et al. (2010) and Gordon et al. (2015) in their work. So for the annotation of the source and target domains in our work, we combined the domains provided by them and provided the same to our annotators.

A number of works in psycholinguistics have explored how metaphors are used in framing different conversational domains (Semino et al., 2018; Cornelissen et al., 2011; Luokkanen et al., 2014; Joris et al., 2014). In the political domain specifically, Brugman et al. (2017) and Boeynaems et al. (2017) explore the role of metaphorical framing for persuasion.

Our work complements existing resources that brings together metaphors with persuasive effects, as stated in Section 1. Previous NLP research in this context has studied how cognitive traits of readers such as personality and prior beliefs impact persuasive effects (Lukin et al., 2017; Durmus and Cardie, 2018). Others looked at the interplay of the characteristics of debaters in persuasive argumentation (Al Khatib et al., 2020). For editorials,

Effect	Liberal	Conservative
No effect	70	201
Reinforcing	131	68
Challenging	99	31
Total	300	300

Table 1: Distribution of persuasive effects over the two political idelogies (liberal, conservative) in our corpus.

El Baff et al. (2020a) studied correlations between the impact of arguments and the the reader's personality and political ideology. Since linguistic style is largely encapsulated by the usage of metaphorical language, our analysis draws comparisons from the work of El Baff et al. (2020b) who study the importance of the writing style of news editorials for the persuasion of readers with different ideologies.

3 Data

This section presents the construction of our news editorial corpus annotated for metaphors. We start with describing the source data, before explaining the annotation task and guidelines. Then, we describe the annotation process, consisting of two pilot studies and one main annotation phase.

3.1 Source Data

As a basis for annotating metaphors, we use the publicly available Webis-News-Editorials18 dataset (El Baff et al., 2018) [1]. This corpus contains ideology-specific annotations of the perceived persuasive effect of 1000 New York Times editorials. In particular, three liberals and three conservative readers each annotated all editorials by reporting whether the editorial *reinforced* their stance by helping them argue better, *challenged* their stance by making them rethink it, or had *no effect* on them.

For our corpus, we randomly sampled 300 of the editorials. They have a mean length of 7.3 paragraphs per editorial. The total number of tokens in the sample is 161 132, that is, 537.1 tokens per editorial on average. The distribution of the frequency of the persuasiveness effects for the liberal and conservative ideologies are shown in Table 1.

3.2 Annotation Task

As outlined in Section 2, past research has predominantly involved the identification of metaphorical usage for single words. However, our inspection

[1] https://webis.de/data/webis-editorial-quality-18.html

of a sample of the corpus revealed that multi-word metaphorical expressions are present in abundance. Hence, we decided to target metaphor identification on two sub-categories: *single metaphor* and *composite metaphor*. Notably, some composite metaphors, in theory, can be also argued to be idioms. However, for the uniformity of our work, we do not sub-categorize metaphorical usage down to idioms in this paper.

To the best of our knowledge, all existing datasets related to metaphors in NLP, restrict their annotation to single-word metaphors. Hence, our work is the first to include multi-word metaphors in the process of curating a dataset involving metaphorical language.

We asked the annotators to read news editorials from the New York Times and to annotate all textual segments (both single-word and multi-word) that encapsulate a utilization of metaphors. In addition, they were asked to identify the source and target domain of the respective metaphor annotated. This led to a two-level annotation task:

- *Level 1.* The first task was to identify all single and composite metaphors. A single metaphor is a one-word metaphor usually represented by a verb, for example, "He is *drowning* in money". It may be an adjective or a noun as well, though, as in "a *nourishing* vacation" or "*nation's* choice", respectively. In contrast, a composite metaphor constructs metaphorical meaning with more than one word, as in this example: "The world has begun *edging away* from the dollar."
- *Level 2.* For each metaphor identified, the annotator was then asked to determine the source and target domains of the metaphor from a set of pre-defined concept (source and target) domains. These domains represent the conceptual mapping of the meaning shift in metaphors, as stated in Section 2.

3.3 Annotation Guidelines

Following from the specified annotation task, our annotation guidelines consisted a *procedure* to identify metaphor as well as source and target domains, along with examples of the annotation task.

Procedure For single metaphors, we relied on the widely-used Metaphor Identification Procedure (MIPVU) by Steen (2010).[2] MIPVU summarizes the following steps to identify metaphors, which we reuse in our work:

- Read the text to get a general understanding of the meaning.
- Determine the lexical units.
- Establish the contextual meaning of the unit.
- Determine if it has a more basic meaning.
- Decide whether the contextual meaning contrasts with the basic meaning but can it be understood in comparison with it.
- If yes, mark the unit as metaphorical.

Composite metaphors consist of a pivot metaphorical word and a context window on the left, right, or both sides of the pivot metaphorical word. For identification, the following steps were devised:

- Find the pivot word, following MIPVU.
- Identify the context window: at least one-word and at most two words on the left, right, or both sides of the pivot word. This window was decided based upon discussing with the annotators after the first two pilot studies, before starting with the main annotations.

Previous work has listed a comprehensive number of source and target domains with respect to different sources of text (Shutova et al., 2010; Gordon et al., 2015). Especially in the work by the latter, one of the prominent sources of text in their dataset are news platforms. We combined the source and target domains of these two works and provided them to the annotators. We also kept the option open for the annotators to add a new source and/or target domain if they found the list we provided to contain insufficient number of candidates for source and target domains.

3.4 Annotation Tool

For our pilot studies and the main annotation, we employed Label Studio, an annotation platform designed to host and support crowdsourced annotations.[3] We customized the platform such that, for each article, an annotator can read the editorial and mark a span of text as single or composite metaphor simply by clicking either of the two buttons we provided in our customized tool. The annotators can then choose those source and target domains from the given set of domains that they deem to be the

[2] Metaphor Identification Procedure, http://www.vismet.org/metcor/documentation/MIPVU.html

[3] LabelStudio, https://labelstud.io

most suitable candidate for the given metaphorical expression in that sentential context.

3.5 Annotation Process

Given the complexity of the task, we carried out two phases of pilot studies with two annotators – both of them experts on the subject domain with previous knowledge about metaphorical meaning construction – before we proceeded to the main annotation study. The annotators were handed the annotation guidelines and briefed about the annotation procedure based on the guidelines.

Pilot Studies Each phase of the pilot study had three randomly sampled editorials from the original corpus. We calculated the agreement between the annotators based on span overlaps as follows:

- For single metaphors, we considered the annotators to agree, if there was an exact match of spans.
- For composite metaphors, we considered them to agree, if there was a full or a partial overlap of spans.

The first phase yielded an observed agreement of 66.7% for the identification of metaphors. We discussed conflicts that arose over the first pilot study with the annotators. Example disagreement covered cases such as whether to consider personification of institutions (e.g., "government") as metaphors or whether the context of composite metaphors can span over more than two tokens. Additionally, the pilot study resulted in adding more frequent source and target domains, such as *people*. The second phase of the pilot study resulted in an observed agreement of 65.0%. Upon another discussion with the annotators, we added *object* to the list of both source and target domains and finalized the guidelines.

Main Study For the main annotation, we hired only one additional annotator who was trained on the revised guidelines. All three annotators then worked with the finalized version of the annotation guidelines [4].

For the final annotations we calculated chance-corrected inter-annotator agreement and consolidated the final annotation in the following process:

1. For each editorial, we converted each annotation into BIO format (Ye et al., 2019), resulting in each token in the text having one of the following labels: B_{single} for single metaphors, $B_{composite}$ and $I_{composite}$ for composite metaphors, and O if the token does not belong to any metaphor.
2. For each editorial, we calculated the pairwise average agreement in terms of Fleiss' κ based on the tokens' BIO tags of the three annotators (annotators 1 & 2, 2 & 3, and 1 & 3).
3. As final inter-annotator agreement, we report the average of the pairwise agreements of the three annotators over all the editorials below.
4. To infer the final annotation of each editorial, we take the annotations of the most reliable annotator. We define this annotator to be the one with the highest agreement with the other two on the corresponding editorial.

As a result, the average agreement scores of Annotators 1, 2, and 3 across all editorials were *0.41*, *0.41*, and *0.40* respectively. This moderate agreement reflects the subjectivity of this task, while underlining the consistency of the annotation and the uniform understanding of the guidelines by the annotators. Additionally, the majority agreement between the three annotators was *0.79*, which shows that the consistency of the absolute agreement is maintained from the pilot studies. Detailed statistics of the corpus are shown in Table 2.

4 Dataset Analysis

In this section, we analyze how metaphor types (single or composite) and domains (source and target) correlate to editorials' effect on readers with different ideologies by providing a numerical analysis. The dataset contains 12,006 metaphors across 300 news editorials, 70% of which are single and 30% composite metaphors (Table 2). El Baff et al. (2018) has additional annotations of three different topics - *global*, *national*, and *state* pertaining to the scope of the news editorials. Table 3 shows the count of single and composite metaphors based on the topic distribution of the editorials.

In the following sections, we apply the same methodology to identify significant differences between the following pairs of effects: Challenging vs. Reinforcing, Challenging vs. No Effect, or Reinforcing vs. No Effect. To achieve this, we initially conduct significance tests on a dependent

[4] Code and data is available here: https://github.com/webis-de/NAACL-24/blob/main

	Conservative						Liberal					
	All metaphors		Single		Composite		All metaphors		Single		Composite	
Effect	**count**	**ratio (std)**	**count**	**ratio (std)**	**count**	**ratio (std)**	**count**	**ratio (std)**	**count**	**ratio (std)**	**count**	**ratio (std)**
Challenging	1,349	0.13 (± 0.04)	928	0.06 (± 0.02)	421	0.08 (±0.04)	4,119	0.13 (± 0.04)	2,875	0.06 (± 0.02)	1,244	0.07 (± 0.03)
No effect	7,887	0.12	5,478	0.05	2,409	0.07	2,404	0.10	1,592	0.04	812	0.06
Reinforcing	2,770	0.12	1,947	0.05	823	0.07	5,483	0.13	3,886	0.06	1,597	0.07
All Effects	**12,006**		**8,353**		**3,653**		**12,006**		**8,353**		**3,653**	

Table 2: Total *count* and mean *ratio* of metaphorical tokens among all news editorials of *all metaphors*, *single*, and *composite* metaphors for *conservative* and *liberal readers* per effect (*challenging*, *no effect* and *reinforcing*). Ratios that are significantly different across the three effects are reported with p-value < 0.05 and 0.001.

Topic	**# Single**	**# Composite**	**# Editorials**
Global	2220	973	79
National	3455	1536	124
State	2678	1144	97
Total	8353	3653	300

Table 3: Count of single and composite metaphors over three topics (global, national, and state) in our corpus.

variable, namely the metaphor ratio in Section 4.1 and the domain count in Section 4.2, across the three editorials' effects (e.g., Challenging) to determine if there are any significant variations among these values. We employ the Anova test (if homogeneity and normality are met) or Kruskal (otherwise). Subsequently, if p < 0.05, we perform posthoc analysis, employing an independent t-test when normality is satisfied and Mann-Whitney otherwise, with Bonferroni correction for each effect pair. Additionally, we calculate the effect size r to quantify the strength of the observed significant differences.

4.1 Metaphorical Effect on Ideology

Table 2 shows the count of metaphors for Conservative (left) and Liberal (right) readers for each effect (*Challenging*, *Reinforcing* and *No Effect*). As mentioned earlier, the distribution of the editorials over the effect is imbalanced for both ideologies. For conservatives, the editorials are distributed with 10%, 67%, and 23% for challenging, no effect, and reinforcing, and for liberals, 33%, 23%, and 44%. Simultaneously, for Conservatives, the number of metaphors in ineffective editorials is higher (7,887) than the effective ones (*Challenging*: 1,349, *Reinforcing*: 2.770). Meanwhile, for liberals, effective editorials have a higher (4,119 and 5,483) number of metaphors than ineffective (2,404). A similar distribution holds for single and composite metaphors.

Additionally, we report the mean of the ratio of metaphorical tokens to total tokens (**ratio**), which ranges between 0.10 and 0.13 when considering both metaphor types and between 0.04 and 0.07 when considering single or composite (Table 2). We conduct the significance test explained earlier on the **ratio**.

Conservatives. We do not observe any significant differences across the three effects, suggesting that metaphors have no significant effect on conservatives. El Baff et al. (2020c) showed, using the same dataset, that conservative readers are not affected by the style features of an editorial but rather by the content, deducing that conservative readers are resistant to the style of a liberal newspaper (The New York Times). Our results complement previous findings since metaphors are also considered style features.

Liberals. On the contrary, Liberal readers are affected by metaphors. Overall, there is a significant difference across the three effects at $p < 0.001$, where Reinforcing and Challenging editorials have significantly higher metaphors than ineffective ones with a medium effect size of $r = .35$ and .29, respectively. The same observations hold for single metaphors (effect size $r = .34$ and $r = .33$ respectively). Whereas, for composite metaphors, the significance is weaker at $p < .05$, and only Reinforcing editorials have more composite metaphors than ineffective ones with an effect size of $r = .2$.

These findings also align with El Baff et al. (2020c), showing that liberals are affected by style features, such as emotional tone and argumentative lexicon. Here, we reinforce previous findings using another style feature: metaphors.

Category	All unique	All count	Source unique	Source count	Target unique	Target count	Domain Examples
Nature	16	**4,466**	**14**	2,256	4	2210	stone, body of water, abyss, nature, geographic feature
Morality and justice	**19**	3,946	9	331	13	**3,615**	law, freedom, crime, duty, right, democracy
Darkness and light	9	3,185	7	**3,136**	4	49	weakness, darkness, moment in time, death, light
Engineering and business	15	2,980	5	702	10	2,278	energy, development, development, competition
Systematic explanations	12	2,952	8	1,625	7	1,327	forceful extraction, position, problem, object
Life cycle and relations	13	2,434	7	816	7	1618	social system, people, family, change, relationship
Health and safety	15	1,204	**14**	1,155	2	49	natural physical force, illness, hazardous
Embodied experience	3	1,014	**1**	788	2	226	emotional experience, emotion, feeling
Journey	7	766	4	307	3	459	past, struggle, story, service, success, journey, time
Animals	9	404	6	395	3	9	animal, addiction, monster, parasite, game
Power and control	4	350	**1**	45	3	305	rule, election, leader, enonomic system
Conflict	**2**	303	2	303	**0**	**0**	war, barrier
Plants	5	261	5	261	**0**	**0**	food, resource, crop, plant
High and low	**2**	**28**	**1**	**27**	1	1	temperature, hurdle

Table 4: The unique and total count of source and target domains over the 14 categories (Gordon et al., 2015) along with examples from the domains selected from our dataset. Highest and lowest numbers are **bolded**.

4.2 Domain Effect on Ideology

The source and target domains contain 131 unique overlapping values: 84 and 59, respectively. Given our dataset's small amount (300), gaining insights into the source and target domains concerning the editorial effect is challenging. For that, we categorize the domains into a systemic taxonomy of ontological categories, and then we conduct our analysis of the relationship between domain category and editorials' effect per ideology.

Domain Categorization

To categorize source and target domain, we rely on the work of Gordon et al. (2015), where they defined and categorized source domains into 14 ontological categories as shown in Table 4, which semantically and conceptually abstract similar source domains. For example, the source domains *body of water* and *abyss* are categorized under *nature*, while *animal* and *monster* are categorized under *animal*, as shown in Table 4.

To categorize the domains (source and target) in our dataset, each goes through the following pipeline: (1) preprocessing (we manually fix the typos), (2) measuring the relatedness between the domain and each of the fourteen categories using Dor et al. (2018)'s Term Relater and (2) assigning the category with the highest score to the domain.

Table 5 shows the distribution of the unique and the total count of domains across the 14 categories where *Morality and Justice* have the highest unique domains (19) and *Conflict* and *High and Low* have the lowest unique number (2).

Darkness and light (e.g., death, light) has the highest encounter (3,136) in the source domain, whereas *Morality and justice* (e.g., law, crime) is the highest for the target domain. Gordon et al. (2015) states that there are no theoretical or practical limitations on the target domains that metaphors can describe, which makes it challenging to categorize them. However, source domains (even though not restricted) are drawn by a common set of familiar scenarios, pinpointing that they are easier to categorize. This explains the difference in the distribution between the categories for source and target; however, we observe some similarities, such as the prominence of the *Nature* (e.g., stone) category in both.

Domain Category Effect on Ideology

Now that each source and target domain is mapped to one of the 14 categories, we conduct our analysis to gain insights into the connection between domain categories and editorials' effect on both ideologies. Table 5 shows the distribution of the categories within each domain across editorials' effect (Challenging, No Effect, and Reinforcing) for each ideology (Conservatives on top and Liberals at the bottom row.).

In addition, we conduct the same analysis as before to calculate significance across effects for the count of each category. As a result, Figure 2 shows the effect size r for each category (y-axis) that had at least a significant difference across a pair of effects (x-axis, e.g., Challenging vs. No Effect)

Effect	Darkness & Light		Nature		Journey		Embodied Experience		Power & Control		Engineering & Business		Life Cycle & Relations		Other	
	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target	Source	Target
Conservative Readers																
Challenging	336	3	275	256	18	41	90	35	5	8	78	219	111	181	447	619
No Effect	2126	30	1454	1,335	183	302	523	154	30	257	459	1613	502	1047	2711	3244
Reinforcing	674	16	527	619	106	116	175	37	10	40	165	446	203	390	939	1138
Liberal Readers																
Challenging	1,010	19	802	75	124	157	265	72	15	48	243	665	314	571	2,772	3,514
No Effect	509	15	331	406	66	78	134	48	3	97	178	461	189	296	2,004	2,028
Reinforcing	1,617	15	1,123	929	117	224	389	106	27	160	281	1,152	313	751	3418	4,460
All Effects	**3,136**	**49**	**2,256**	**2,210**	**307**	**459**	**788**	**226**	**45**	**305**	**702**	**2,278**	**816**	**1,618**	**8,194**	**10,002**

Table 5: The count of source and target domain categories, per editorial effect (*Challenging*, *No Effect* and *Reinforcing*) with the total of each category-source and -domain (*All Effects*). Seven out of 14 categories are shown and the other 7 categories are combined under *Other*. The counts that are significantly different across effects are reported with p_value < 0.05 , 0.01 and 0.001.

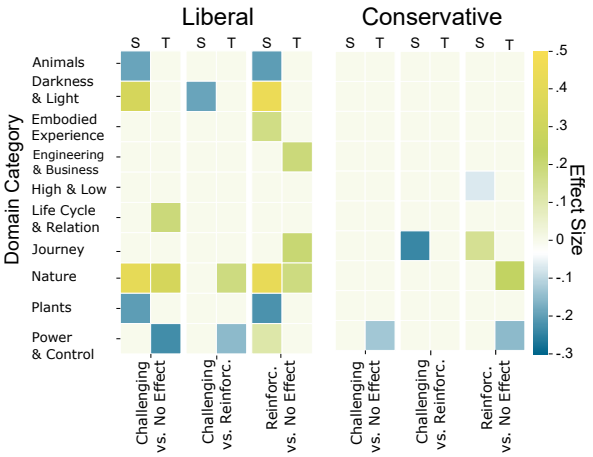

Figure 2: Heatmap for each domain category for Liberal readers (left) and Conservative readers (right). Each ideology has a source (*S*) and target (*T*) heatmap. The y-axis represents each category, and the x-axis represents each effect-pair (e1 vs e2). Each cube represents the effect size r: a green and yellow color indicates a positive effect size where e1 has a significantly higher number than e2 (e1 » e2), whereas a blue color indicates the opposite (e1 « e2).

for either ideology. Each cell is colour-coded for each effect pair (effect1 vs. effect2), where blue indicates that the count of a category in effect2 is significantly higher than for effect1.

The opposite interpretation stands for green and yellow. In general, the results coincide with our previous observations. For liberals, we observe richer results where nine categories significantly differ across editorials' effects compared to four categories for conservatives (Figure 2).

Our findings are in accordance with the definition provided by El Baff et al. (2018) regarding effective editorials that can have two values: *Challenging* if the editorial holds a stance **opposite** to the reader, or *Reinforcing* if the editorial holds the **same** stance as the reader. Our key findings are described below, based on Figure 2.

Conservatives. The source domains categorized under Journey are significantly higher for reinforcing editorials than others, implying that this source category is effective when the reader shares the same stance as the editorial ($r = .24$ and $r = .15$ vs. Challenging and No Effect). Additionally, reinforcing editorials have fewer source domains with High & Low categories compared to ineffective ones with a very weak effect size of $r = .07$. For target domains under the Power & Control category, effective editorials have lower counts than ineffective ones, with a strong effect size of $r = .13$ and $r = .15$ for Challenging and Reinforcing, respectively.

Liberals. For the source domains, effective editorials have significantly fewer domains categorized under Animals and Plants with effect sizes $r = .19 - .22$. The opposite holds for Darkness & Light and Nature with a strong effect size ($r = .33$ and $r = .44$ vs. Challenging and Reinforcing, respectively). Also, for Reinforcing editorials, Power & Control ($r = .11$) and Embodied Experience ($r = .17$) seem more prominent than ineffective editorials.

For the target domain, a similar observation holds for the Nature category, where it is more prominent in effective editorials ($r = .33$ and $r = .18$ for Challenging and Reinforcing, respectively). Categories such as Engineering & Business and Journey are more prominent in Reinforcing editorials than ineffective ones ($r = .2$). Finally, challenging editorials have significantly fewer Power & Control target domains compared to ineffective ($r = .23$) and reinforcing ($r = .15$) editorials, implying that this target category is ineffective when the reader has an opposite stance to the editorial's stance.

5 Conclusion

The framing of political discourses often relies heavily on the use of metaphors. In this paper, we introduce a dataset which includes two levels of annotations in the intersection of metaphorical usage and the persuasive effects of political ideologies on readers. The first level is designed to identify metaphors, while the second level aims to determine the domains from which these metaphors originate (source domain) and to which they are applied (target domain).

Our findings show that liberal readers are affected significantly by metaphors, whereas conservatives are more resistant to them. The impact of metaphors on ideologies varies based on their conceptual domains, influencing how political opinions are either challenged or reinforced in news editorials. For example, Liberals are affected by metaphors in the Darkness & Light (e.g., death) source domains, whereas the source domain of Nature affects conservatives more significantly.

6 Acknowledgment

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number TRR 318/1 2021 – 438445824. We thank the anonymous reviewers for their helpful feedback.

7 Limitations

The usage of metaphors in language is grounded, which means that some words and phrases can be seen to be both metaphorical and not metaphorical depending on the subjectivity of the context in language, such as for "inflation is going *higher*". In our work, we consider all degrees of metaphoricity as metaphorical usage, and for simplicity do not distinguish between the various degrees.

However, there may still be metaphorical utterances that we missed in our annotation process. In particular, to foster a uniform metaphor handling among the annotators, we confined the context window of composite metaphors to be at most two words on either end of the pivot word. That being said, there may be cases where the entire sentence would be needed as the context window in order to properly identify a metaphor.

Finally, we point out that our analysis of the interactions between metaphors and persuasive effects are limited by the expressiveness of the human judgments. Persuasion is an intrinsically subjective concept that is not only affected by political ideologies. Hence, we may have observed correlations that confounded by other characteristics of the human annotators. Since the persuasive effect annotations came from previous work, we had no way to further control for this, but future work should validate our results in comparison studies.

8 Ethical Statement

We consider no conceivable immediate potential ethical issues or threat to be caused by our corpus, since we only analyzed semantic concepts in an existing corpus. Each of the annotators was paid $ 12.50 per hour, which is in line with the standards of fair payment of the host institutions of all authors of this paper. We consider no potential threat to be caused by our dataset.

References

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. 2020. Exploiting personal characteristics of debaters for predicting persuasiveness. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7067–7072, Online. Association for Computational Linguistics.

Ramy Baly, Giovanni Da San Martino, James Glass, and Preslav Nakov. 2020. We can detect your bias: Predicting the political ideology of news articles. *arXiv preprint arXiv:2010.05338*.

Felix Biessmann, Pola Lehmann, Daniel Kirsch, and Sebastian Schelter. 2016. Predicting political party affiliation from text. *PolText*, 14(14):2016.

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.

Britta C Brugman, Christian Burgers, and Gerard J Steen. 2017. Recategorizing political frames: A systematic review of metaphorical framing in experiments on political communication. *Annals of the International Communication Association*, 41(2):181–197.

Pere-Lluís Huguet Cabot, Verna Dankers, David Abadi, Agneta Fischer, and Ekaterina Shutova. 2020. The pragmatics behind politics: Modelling metaphor, framing and emotion in political discourse. In *Findings of the association for computational linguistics: emnlp 2020*, pages 4479–4488.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Tuhin Chakrabarty, Xurui Zhang, Smaranda Muresan, and Nanyun Peng. 2021. Mermaid: Metaphor generation with symbolism and discriminative decoding. *arXiv preprint arXiv:2103.06779*.

Jonathan Charteris-Black. 2009. Metaphor and political communication. In *Metaphor and discourse*, pages 97–115. Springer.

Paul Chilton and Mikhail Ilyin. 1993. Metaphor in political discourse: The case of thecommon european house'. *Discourse & society*, 4(1):7–31.

Joep P Cornelissen, Robin Holt, and Mike Zundel. 2011. The role of analogy and metaphor in the framing and legitimization of strategic change. *Organization Studies*, 32(12):1701–1716.

Erik-Lân Do Dinh, Hannah Wieland, and Iryna Gurevych. 2018. Weeding out conventionalized metaphors: A corpus of novel metaphor annotations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1424.

Liat Ein Dor, Alon Halfon, Yoav Kantor, Ran Levy, Yosi Mass, Ruty Rinott, Eyal Shnarch, and Noam Slonim. 2018. Semantic relatedness of wikipedia concepts–benchmark data and a working solution. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Esin Durmus and Claire Cardie. 2018. Exploring the role of prior beliefs for argument persuasion. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1035–1045, New Orleans, Louisiana. Association for Computational Linguistics.

Roxanne El Baff, Khalid Al Khatib, Benno Stein, and Henning Wachsmuth. 2020a. Persuasiveness of news editorials depending on ideology and personality. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 29–40, Barcelona, Spain (Online). Association for Computational Linguistics.

Roxanne El Baff, Henning Wachsmuth, Khalid Al-Khatib, and Benno Stein. 2018. Challenge or Empower: Revisiting Argumentation Quality in a News Editorial Corpus. In *22nd Conference on Computational Natural Language Learning (CoNLL 2018)*, pages 454–464. Association for Computational Linguistics.

Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2020b. Analyzing the Persuasive Effect of Style in News Editorial Argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3154–3160, Online. Association for Computational Linguistics.

Roxanne El Baff, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2020c. Analyzing the persuasive effect of style in news editorial argumentation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3154–3160.

Farahman Farrokhi and Sanaz Nazemi. 2015. The rhetoric of newspaper editorials. *International Journal on Studies in English Language and Literature (IJSELL)*, 3(2):155–161.

Hongyu Gong, Kshitij Gupta, Akriti Jain, and Suma Bhat. 2020. Illinimet: Illinois system for metaphor detection with contextual and linguistic information. In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 146–153.

Jonathan Gordon, Jerry R 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.

Willem Joris, Leen d'Haenens, and Baldwin Van Gorp. 2014. The euro crisis in metaphors and frames: Focus on the press in the low countries. *European Journal of Communication*, 29(5):608–617.

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.

George Lakoff. 1995. Metaphor, morality, and politics, or, why conservatives have left liberals in the dust. *Social Research*, pages 177–213.

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

George Lakoff and Mark Johnson. 2020. Conceptual metaphor in everyday language. In *Shaping entrepreneurship research*, pages 475–504. Routledge.

Chang Li and Dan Goldwasser. 2019. Encoding social information with graph convolutional networks forPolitical perspective detection in news media. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2594–2604, Florence, Italy. Association for Computational Linguistics.

Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument Strength is in the Eye of the Beholder: Audience Effects in Persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 742–753. Association for Computational Linguistics.

Matti Luokkanen, Suvi Huttunen, and Mikael Hildén. 2014. Geoengineering, news media and metaphors: Framing the controversial. *Public Understanding of Science*, 23(8):966–981.

Jeffery Scott Mio. 1997. Metaphor and politics. *Metaphor and symbol*, 12(2):113–133.

Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana. 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.

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, Ingrid Scharlau, and Henning Wachsmuth. 2023. Modeling highlighting of metaphors in multitask contrastive learning paradigms. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4636–4659, Singapore. Association for Computational Linguistics.

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.

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.

Gerard Steen. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing.

Enrica Troiano, Laura Oberländer, and Roman Klinger. 2023. Dimensional modeling of emotions in text with appraisal theories: Corpus creation, annotation reliability, and prediction. *Computational Linguistics*, 49(1):1–72.

Wei Ye, Bo Li, Rui Xie, Zhonghao Sheng, Long Chen, and Shikun Zhang. 2019. Exploiting entity bio tag embeddings and multi-task learning for relation extraction with imbalanced data.

Dongyu Zhang, Minghao Zhang, Heting Zhang, Liang Yang, and Hongfei Lin. 2021. Multimet: A multimodal dataset for metaphor understanding. 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 3214–3225.