

# On the Influence of Discourse Relations in Persuasive Texts

Nawar Turk<sup>†,\*</sup>, Sevag Kaspar<sup>†</sup>, Leila Kosseim<sup>†</sup>

<sup>†</sup> Dept. of Computer Science and Software Engineering, Concordia University

Published in: *Proceedings of the 38th Canadian Conference on Artificial Intelligence (CanAI 2025)*, Calgary, Alberta, May 26–27, 2025.

Available at: <https://caiac.pubpub.org/canadian-ai-2025-long-papers>

## Abstract

This paper investigates the relationship between Persuasion Techniques (PTs) and Discourse Relations (DRs) by leveraging Large Language Models (LLMs) and prompt engineering. Since no dataset annotated with both PTs and DRs exists, we took the SemEval 2023 Task 3 dataset labelled with 19 PTs as a starting point and developed LLM-based classifiers to label each instance of the dataset with one of the 22 PDTB 3.0 level-2 DRs. In total, four LLMs were evaluated using 10 different prompts, resulting in 40 unique DR classifiers. Ensemble models using different majority-pooling strategies were used to create 5 silver datasets of instances labelled with both persuasion techniques and level-2 PDTB senses. The silver dataset sizes vary from 1,281 instances to 204 instances, depending on the majority pooling technique used. Statistical analysis of these silver datasets shows that six discourse relations (namely CAUSE, PURPOSE, CONTRAST, CAUSE+BELIEF, CONCESSION, and CONDITION) play a crucial role in persuasive texts, especially in the use of *Loaded Language*, *Exaggeration/Minimisation*, *Repetition* and to cast *Doubt*. This insight can contribute to detecting online propaganda and misinformation, as well as our general understanding of effective communication.

All code and data developed for this work are publicly available at <https://github.com/CLaC-Lab/Persuasion-Discourse-Mapping>.

**Keywords:** Discourse Relations, Persuasion Techniques, Large Language Models

## 1. Introduction

Persuasion techniques play a crucial role in communication and are often used to spread propaganda. Due to their significance in shaping public opinion and today’s ease at generating text via LLMs, a growing body of work in Natural Language Processing (NLP) has studied persuasive texts computationally and has developed inventories of specific writing techniques such as *Appeal to Authority* and *Loaded Language* to create more persuasive narratives.

On the other hand, the field of computational discourse analysis has a long history of studying the coherence structure of discourse by characterizing how text spans are connected logically to make a text coherent. In this view, discourse relations can be seen as general in nature as they characterize texts with varying communicative goals, whereas persuasion techniques characterize the specific relations used for the purpose of persuasion.

This paper investigates the relationship between PTs and DRs. To do so, we first developed LLM-based discourse relation classifiers capable of labelling text spans with PDTB 3.0 level-2 senses. We then applied ensembles of the top classifiers to label the human-labelled PT SemEval 2023 Task 3 English corpus ([1]) to create a series of 5 silver datasets with both DR and PT annotations. These silver corpora were then mined to uncover patterns between discourse relations and persuasive language.

The main contributions of this paper include:

- (1) A comparative evaluation of various LLM-based models for annotating level-2 PDTB senses.

\* nawar.turk@mail.concordia.ca

- (2) The design and release of silver corpora with both persuasion techniques and discourse relations annotations.
- (3) The identification of the discourse relations that are more often used to implement specific persuasion techniques in texts.

This insight can help improve both the automatic identification of persuasive techniques and our broader understanding of effective communication to help combat online propaganda and misinformation.

## 2. Background

Persuasion techniques are essential for communication, but are often used to disseminate online propaganda. Driven by the increasing ease of text generation through LLMs, a growing body of work in NLP research has addressed the study of persuasion techniques, especially in the context of automatic detection of propaganda (e.g. [1–4]). Early work focused solely on binary document-level classification (propaganda or not), then [5] extended the classification to 4 specific types (*trusted*, *satire*, *hoax*, and *propaganda*). Later models addressed the detection of more fine-grained persuasion techniques within shorter textual spans (e.g. [1–4]). Recent work has also extended the detection of persuasion techniques to multimodal media such as memes [4, 6]. Parallel to this, several inventories of persuasion techniques have been developed based on rhetorical and psychological techniques. Such inventories include the 6-label fallacy inventory of [7], the European Commission inventory [8], and the related inventory of [9]. This latter inventory is currently the most widely used as it was adopted by the SemEval 2021, 2023 and 2024 Shared Tasks on the detection of persuasion techniques [1, 4]. This list defines around 20 persuasion techniques such as *causal oversimplification*, *name calling* and *smear*. Example (1) shows a sentence labelled with a *causal oversimplification* from the SemEval 2023 Task 3 definition<sup>1</sup>.

- (1) `"text": The reason New Orleans was hit so hard with the hurricane was because of all the immoral people who live there.`  
`"PT": causal oversimplification`

Similar to persuasion techniques, discourse relations characterize the connection between text spans. However, discourse relations focus on the more general structure of a text by indicating how text spans are connected logically to make a text coherent. Discourse analysis has a longer history of research in the field of NLP, and several linguistic frameworks have been proposed to model discourse relations<sup>2</sup>. The most widely used frameworks are the Rhetorical Structure Theory (RST) [10] and the Penn Discourse Treebank (PDTB) [11, 12]. Following these frameworks, several annotated corpora have been developed, such as the RST-DT [13] and 3 versions of the Penn Discourse TreeBank (PDTB) [12, 14, 15]. The PDTB takes into account discourse connectives (e.g. *but*, *because*, ...) and labels these cues with a discourse relation, called discourse *sense*. The most recent version of the PDTB (PDTB 3.0 [15]) organizes these senses in a 3-tier hierarchy: level-1 which includes 4 basic senses (e.g. TEMPORAL, CONTINGENCY), level-2 which includes 22 more specific senses (e.g. CAUSE, CONDITION, PURPOSE), and level-3 which includes very fine-grained senses. Most work in the literature has used either level-1 or level-2 senses. Using the PDTB level-2 senses, Example (1) would be annotated with a CAUSE discourse relation as in Example (2).

- (2) `"text": The reason New Orleans was hit so hard with the hurricane was because of all the immoral people who live there.`  
`"DR": Cause`

<sup>1</sup><https://propaganda.math.unipd.it/semeval2024task4/definitions.html>

<sup>2</sup>Discourse relations are also known as “coherence relations” or “discourse senses”.

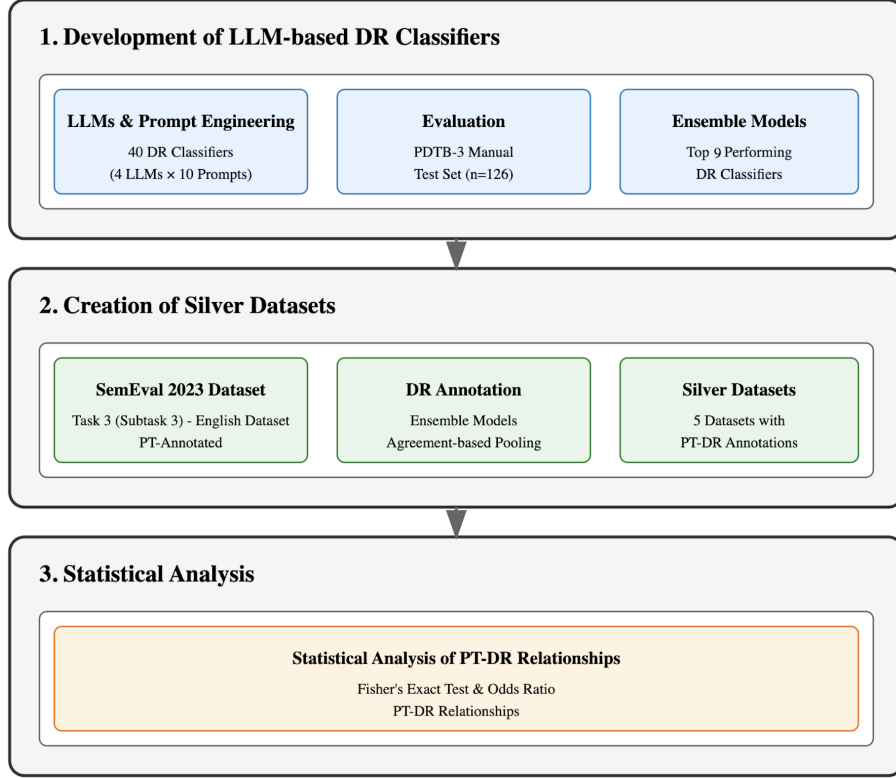


Figure 1. Overall Methodology

Recent work has investigated the interplay between discourse structures and persuasive language. [16, 17] have examined the relationship between discourse relations and propaganda. [16] explored the connection between the four top-level PDTB discourse relations (COMPARISON, CONTINGENCY, TEMPORAL, and EXPANSION) and the eight subtypes of news discourse structure defined in [18] with the presence of propaganda. Similarly, [17] examined the correlation between persuasion techniques and RST rhetorical relations. Additionally, [19] investigated the distribution of discourse connectives and 3 PDTB level-2 relations across various persuasive texts (talks, interviews, and articles); however, no specific persuasion technique was considered. To complement previous work, the goal of our paper is to systematically investigate the mapping between all level-2 PDTB discourse relations, and all persuasion techniques found in the SemEval 2023 dataset.

### 3. Overall Methodology

To explore the relationship between persuasion techniques (PTs) and discourse relations (DRs), we needed a dataset annotated with both types of information. Since no such dataset was found, we created a series of 5 silver datasets. We chose the SemEval 2023 Task 3 English dataset [1] that has been manually labelled with PTs as a starting point and annotated it automatically with DRs. We chose to automatically label DRs as opposed to starting with a DR-annotated dataset and labelling it with PTs for a few reasons. First, since significantly more research has been done on discourse parsing and PDTB than on persuasion techniques, we expected LLMs to be more knowledgeable about identifying PDTB senses than about identifying PTs. In addition, because the PDTB senses are organized hierarchically, this

allowed us to label instances at different levels of granularity should the DR labelling be ambiguous.

To annotate the SemEval 2023 Task 3 dataset, we considered using open-source PDTB parsers (e.g. [20] and [21]); however, we experimented with LLMs and prompt engineering given their success in a variety of NLP-related tasks, and more recently in discourse parsing ([22, 23]).

As shown in Figure 1, the overall methodology consists of 3 main steps:

**Development of LLM-based DR Classifiers:** First, we developed discourse relation classifiers using LLMs and prompt engineering. We experimented with 4 LLMs and 10 prompts for a total of 40 configurations. After evaluating them on a manually-annotated PDTB test set, we kept the top 9 best-performing classifiers to create an ensemble model. This is described in Section 4.

**Creation of Silver Datasets:** The top-performing DR classifiers were then used to annotate the SemEval 2023 Task 3 [1] with discourse relations. Using various agreement-based pooling strategies, we created 5 silver datasets annotated with both DRs and PTs with varying sizes and annotation confidence levels. This is described in Section 5.

**Statistical Analysis between PTs and DRs:** Finally, the silver datasets were used to conduct statistical analyses to identify the relationship between persuasion techniques and discourse relations. Our analysis revealed both strong associations between specific PTs and DRs, particularly with causal and contrastive relations. This is described in Section 6.

## 4. LLM-based Discourse Relations Classifiers

The task of the LLM-based DR classifiers is to predict one of the 22 level-2 senses from the PDTB 3.0 for each input instance.

### 4.1. Models and Prompt Design

To develop the DR classifiers, we experimented with 4 different LLMs and 10 prompts, resulting in 40 DR classifiers<sup>3</sup>. The four LLMs are `gpt-4o`, `Gemini 1.5-pro`, `Gemini 2.0-flash-exp`, and `Claude` (version 3.5-Sonnet-20241022). Each LLM was queried using 10 prompts developed based on a variety of design elements [24], such as zero-shot instructions, few-shot examples, definitions, and chain-of-thought reasoning. Table 1 shows the description of these prompts. Although all prompts were designed to predict a PDTB level-2 sense, prompts 1 to 5 asked directly to identify a level-2 sense, while prompts 6 to 10 followed a chain-of-thought reasoning, requiring the LLM to first segment the given instance into two discourse arguments, identify the level-3 sense before determining the level-2 parent sense. The prompts also varied with respect to the supporting information they provided, such as whether they provided definitions (at level-2, level-3, or both) and examples (at level-2 or level-3)<sup>4</sup>. Figure 2 shows an example of Prompt 3, which provides the PDTB level-2 senses, their definitions and an example.

#### 4.1.1. Classifier Evaluation

To evaluate the 40 DR classifiers, we used a test set of 126 instances extracted manually from the PDTB 3.0 Annotation Manual [15] with their gold standard senses. The nuances of instances were chosen to include at least one PDTB level-3 sense, ensuring that all level-2

<sup>3</sup>We call a DR classifier the combination of a specific LLM with a specific prompt.

<sup>4</sup>All prompts and their specifications are publicly available at <https://github.com/CLaC-Lab/Persuasion-Discourse-Mapping>.

"instruction":	Analyze the following paragraph and identify the most appropriate PDTB Level 2 discourse relation label from the list below. Use the definitions and examples provided to guide your decision. Only return the label with no additional explanation.
"paragraph":	{paragraph}
	<b>**Available Labels with Definitions and Examples:**</b>
"label":	Synchronous
"definition":	Temporal overlap between events.
"example":	"While the cake was baking, she prepared the icing."
"label":	Asynchronous
"definition":	One event precedes another.
"example":	"First, he packed his bags. Then, he left for the airport."
"label":	Cause
"definition":	Causal influence without a conditional relation.
"example":	"The roads were icy, thus causing the accident."
"label":	Cause+Belief
"definition":	Evidence provided to induce belief.
"example":	"The footprints in the snow led her to conclude someone had been there."
[...]	

Figure 2. Excerpt from Prompt 3 showing four PDTB level 2 senses with definitions and examples. [...] refers to the description of the remaining PDTB level 2 senses.

Prompt ID	Sense level asked for	Definition provided for	Examples provided for	Chain-of-thought reasoning
1	level-2	n/a	n/a	no
2	level-2	level-2	n/a	no
3	level-2	level-2	level-2	no
4	level-2	level-2	level-2	no
5	level-2	level-2, level-3	n/a	no
6	level-3, level-2	level-3	n/a	yes
7	level-3, level-2	level-3	n/a	yes
8	level-3, level-2	level-2, level-3	n/a	yes
9	level-3, level-2	level-2, level-3	n/a	yes
10	level-3, level-2	level-3	level-3	yes

Table 1. Prompt Descriptions for the LLM-Based Discourse Relation Classifiers

senses were covered. To be representative of an actual text, the dataset includes both explicit and implicit relations. We used this dataset as opposed to a larger one to reduce the cost of the classifier evaluation.

To evaluate the classifiers, each LLM+prompt combination was run twice, and the results were averaged. We assessed their performance on the test set using three metrics:

- (1) Macro F1 scores.
- (2) Hallucination rate: the percentage of non-valid level-2 sense predictions. This includes non-PDTB senses, “no relation” responses, and level-1 senses. Level-3 sense predictions were not considered hallucinations as they were mapped to their level-2 parent sense. Note that macro-F1 scores do not include hallucinations in their calculation, as these can easily be filtered out.
- (3) Prediction consistency: the percentage of identical predictions across the two independent runs.

LLM	Macro F1	Hallucinations	Consistency
claude-3-5-sonnet-20241022	0.35	1.6 %	79.9 %
gemini-1.5-pro	0.31	0.7 %	90.9 %
gemini-2.0-flash-exp	0.30	0.1 %	85.6 %
gpt-4o	0.23	0.2 %	71.2 %

Table 2. Performance of the LLMs Across all 10 Prompts on the PDTB Test Set. Results are averaged across all prompts and runs per LLM.

Table 2 shows a summary of the performance of each LLM; while Figure 3 shows details of each classifier. In Figure 3, LLMs are grouped by the prompt they use (1 to 10), and each bar represents the average of the two independent runs. The average for each prompt across all LLMs is shown as a dot (between the orange and green bars). As Table 2 and Figure 3 show, **claude-3-5-sonnet-20241022** achieves the highest average macro F1 score of 0.35. However, this advantage is accompanied by the highest number of hallucinations (1.6%) and moderate prediction consistency (79.9%). In contrast, **gemini-1.5-pro** demonstrates strong overall performance; although its F1 score is slightly lower at 0.31, it exhibits the highest consistency rate of 90.9 % and lower hallucinations, making it a reliable choice for stable performance. **gemini-2.0-flash-exp** also reports a macro F1 score of 0.30, similar to **gemini-1.5-pro**, but distinguishes itself by having the fewest hallucinations (0.1%). Conversely, **gpt-4o**, despite making fewer hallucinations, demonstrates the lowest overall performance. Overall, these findings reveal the trade-offs between accuracy and reliability, but no individual classifier yields a strong enough performance to be used to annotate a dataset automatically.

#### 4.1.2. Ensemble Classifiers

Given that individual classifier performances were below expectations, we adopted an ensemble approach using majority voting. From the 40 classifiers, we selected the nine best-performing ones based on the highest macro F1 scores, strong prediction consistency between runs, and lower hallucination rates. These top-performing classifiers were used as an ensemble by pooling their predictions and taking the majority prediction. By imposing that at least  $n$  classifiers ( $n \in [5...9]$ ) agree on the prediction, we created 5 ensemble classifiers. As shown in Figure 4, the macro F1 scores of these 5 ensemble classifiers increase steadily from 0.45 (for  $n = 5$ ) to 0.74 (for  $n = 9$ ). In addition, given the agreement constraint, none of the ensemble classifiers generated any hallucinations. Given these higher performances, we used these ensemble models to annotate discourse relations in the SemEval 2023 Task 3 (Subtask 3) dataset [1], thereby creating 5 silver datasets.

## 5. Creation of Silver Datasets

To create silver datasets with both persuasion techniques and discourse relations labels, we used the human-annotated SemEval 2023 Task 3 (Subtask 3) [1] dataset annotated with PTs and used the ensemble-models to annotate it automatically with DRs. In order to use this dataset for DR labelling, we had to pre-process it.

Since some instances were labelled with multiple PTs, we split multi-label entries into individual instances, with each instance labelled with a single PT. This pre-processing increased the dataset size from 3,760 to 5,853 instances.

The dataset was extremely imbalanced, particularly for *Loaded Language* and *Name Calling/labelling*, which accounted for 30.9% and 16.7% of the instances, respectively. To reduce the imbalance, we randomly undersampled the over-represented PTs. The reduced

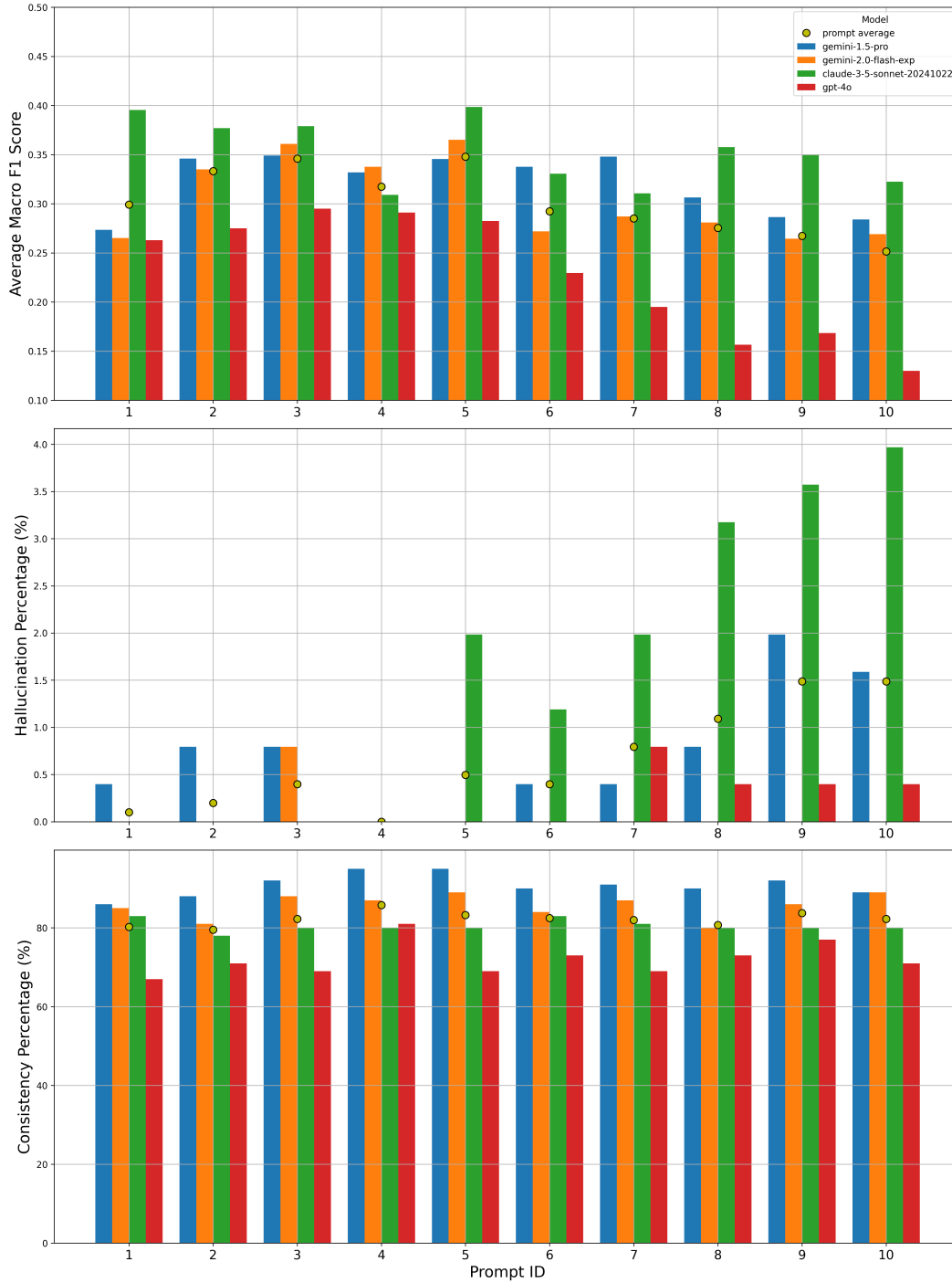


Figure 3. Impact of Prompts on LLM Performance for DR Classification. Each bar represents the average of two independent runs per LLM on that prompt. Prompt average shows the mean performance across all four LLMs for each prompt.

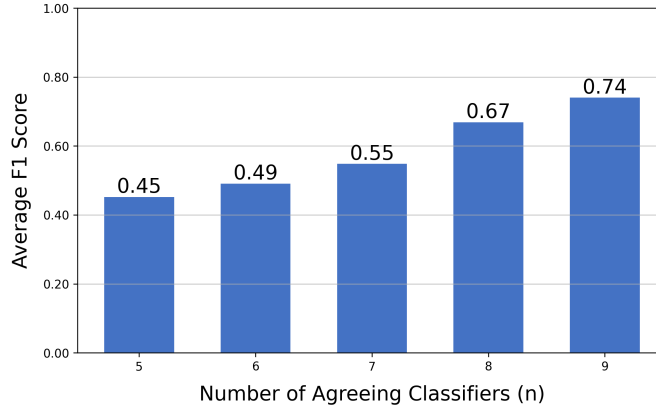


Figure 4. Macro F1 Score of the Ensemble DR Classifiers

"text":	"While we may never know the exact circumstances surrounding what transpired in the shooter's hotel room, the information being released not just by law enforcement, but witnesses to the event who recorded hundreds of cumulative hours of video and audio, now calls the entirety official story into question."
"PT":	"Doubt",
"DR":	"Concession"

Figure 5. Example of a Dually Annotated Instance in a Silver Dataset.

dataset is still imbalanced, but the top 2 PTs account for a more manageable 30.1% of the instances, down from 47.6%.

The next pre-processing step filtered out instances that contained more than 2 sentences. This was done because PDTB parsing typically focuses on identifying relations across arguments within a single sentence or across adjacent sentences. To avoid cases where multiple DRs would be valid labels, we filtered out instances with more than 2 sentences. After this pre-processing step, the dataset was reduced to 2,064 instances.

### 5.1. Silver Datasets

We used the ensemble DR classifiers to annotate the processed SemEval PT dataset with PDTB 3.0 level-2 senses as DRs. As described in Section 4.1.2, we experimented with 5 majority-agreement pooling strategies based on the agreement of  $n$  classifiers out of 9 (where  $n \in [5...9]$ ). The resulting silver datasets ranged in size from 1,281 instances for Silver-5 (i.e.  $n = 5$ ), 937 for Silver-6, 641 for Silver-7, 388 for Silver-8, and 204 for Silver-9. These five silver datasets show the inherent trade-off between dataset size and annotation confidence.

Figure 5 shows a dually annotated instance from the Silver-5 dataset, while Figure 6 shows the distribution of PTs and DRs in the same dataset. As the figure shows, the distribution of DRs in Silver-5 is imbalanced, with some DRs being significantly over-represented, but this imbalance is to be expected and also occurs in the entire PDTB 3.0 and many other discourse labeled corpora [25].

## 6. Statistical Analysis of the Silver Datasets

The goal of the silver datasets was to explore the relationship between persuasion techniques and discourse relations. To have statistically significant results, our analysis focused on PTs and DRs that appear at least 25 times each in the silver datasets.

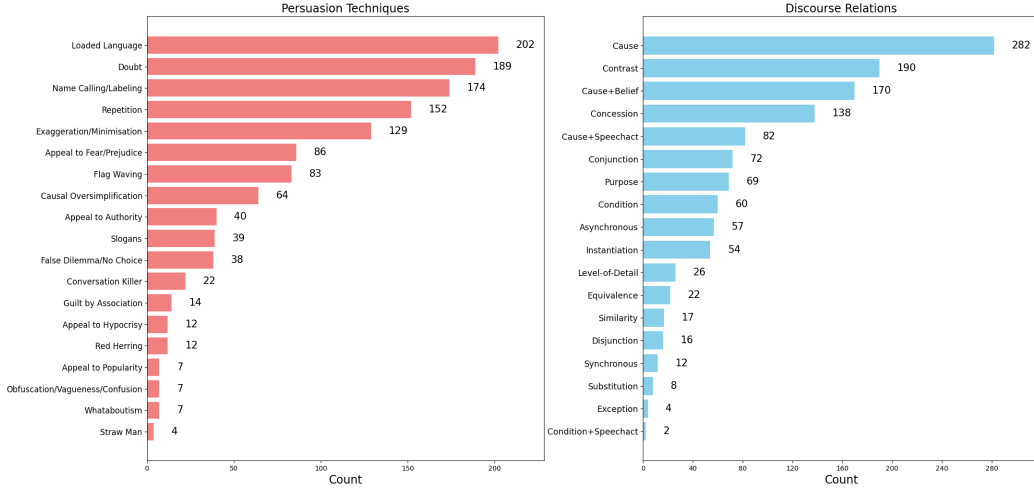


Figure 6. Distribution of PTs and DRs in the Silver-5 Dataset (1,281 instances)

We used Fisher’s exact test to identify statistically significant associations between DRs and PTs. Since some contingency tables of PT-DR pairs contain values less than 5, we used Fisher’s exact test as it is known to be more accurate than the  $\chi^2$  test for small sample sizes. The odds ratio (OR) was then used to determine the direction of these associations, with values greater than 1 indicating positive associations. Figure 7 shows a heat map of all PT-DR pairs and the number of silver datasets where the pair has a statistically significant positive association (i.e. Fisher’s  $p \leq 0.05$  and  $OR > 1$ ). For example, the PT *Loaded Language* has a statistically significant association with a CAUSE DR in all 5 silver datasets, while no dataset identified an association with a CONTRAST DR. As Figure 7 shows, several PT-DR associations clearly stand out. The PDTB senses of CAUSE, PURPOSE, and CONTRAST are the most frequently associated DRs, each paired with five, three, and four PTs, respectively.

Persuasion Techniques such as *Loaded Language* and *Exaggeration/Minimisation* strongly correlate with the CAUSE discourse relation while the *Repetition* PT is consistently expressed through PURPOSE and CONTRAST discourse relations. This suggests that persuasive texts often rely on causal reasoning, purposive statements, and contrasting ideas to have more impact. Similarly, the CAUSE+BELIEF discourse relation shows associations with multiple PTs, including *Doubt* and *Casual Oversimplification*.

Notably, only 6 out of 22 senses showed significant correlations with persuasion techniques. The practical implication is that when writing persuasive text, focusing on these six discourse relations (CAUSE, PURPOSE, CONTRAST, CAUSE+BELIEF, CONCESSION, and CONDITION) would likely lead to more effective persuasion.

## 7. Conclusion and Future Work

This paper explored the relationship between discourse relations (DRs) and persuasion techniques (PTs) through the use of Large Language Models (LLMs). Several individual LLM-based DR classifiers were developed using a variety of prompt engineering methods. Although no individual classifier reached a high enough performance to annotate a dataset with PDTB level-2 senses automatically, ensemble methods reached macro F1 ranging from

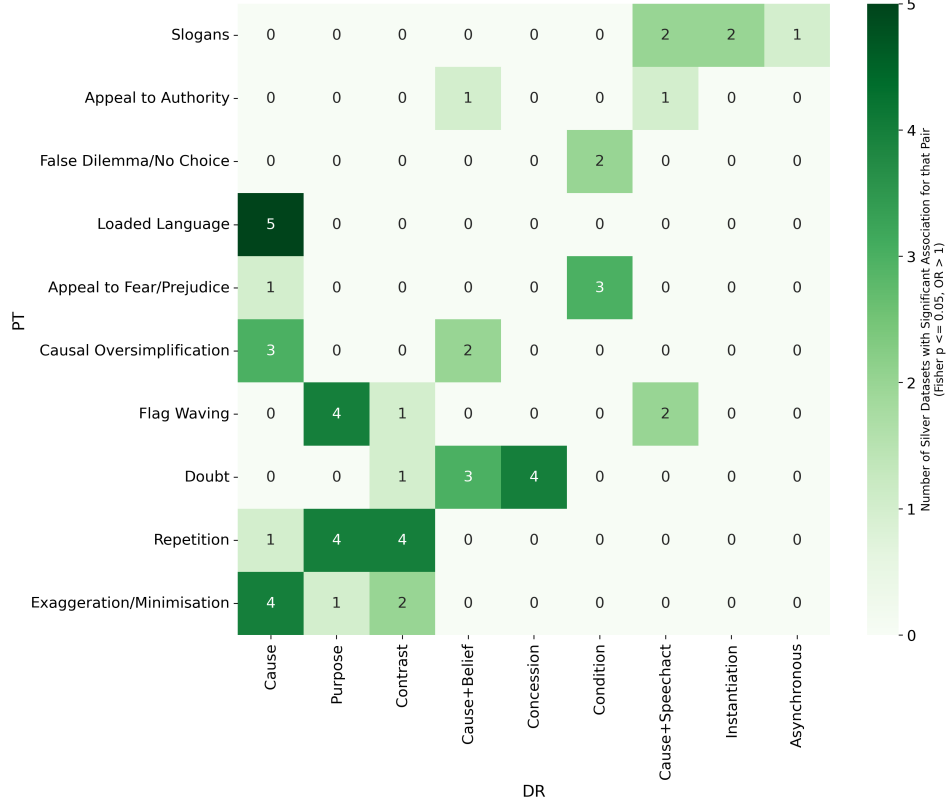


Figure 7. Statistically Significant PT-DR Associations in the Silver Datasets (Fisher’s exact test:  $p \leq 0.05$ ,  $OR > 1$ ).

0.45 to 0.74. These ensemble models were then used to build silver corpora based on the SemEval 2023 Task 3 (Subtask 3) [1] dataset with human-labelled PTs and automatic DRs. A statistical analysis of these silver datasets allowed us to identify pairs of PTs and DRs that co-occur more often, as well as six central discourse relations (CAUSE, PURPOSE, CONTRAST, CAUSE+BELIEF, CONCESSION, and CONDITION) that play a crucial role in persuasive communication. This insight can advance the automatic detection of persuasive techniques and contribute to the detection of online propaganda and misinformation, as well as our general understanding of effective communication.

Future work includes improving the performance of the LLM-based classifiers in order to create larger and more precise silver datasets to allow for a deeper analysis. This work has leveraged a manually-annotated PT corpus (the SemEval 2023 Task 3 dataset) and labelled it with DRs automatically; however, the other direction should also be investigated – i.e. leveraging a manually DR-annotated dataset, such as the PDTB 3.0 [15] with  $\approx 46K$  relations, and building classifiers to label these instances with PTs. Also, this work has assumed that an instance from the dataset can be labelled with a single DR. We should investigate to what degree this assumption holds, as several DRs may be appropriate. Recent work by [26] supports this view by treating implicit discourse relations identification as a multi-label task. Finally, this work has only investigated English texts, yet the SemEval 2023 Task 3 dataset does include annotations for a variety of languages (French, German, Italian, Polish, and Russian). It would be interesting to explore to what extent the relation between PTs and DRs is language-independent or if different languages (or cultures) use discourse relations differently in persuasive communication.

## Acknowledgments

The authors would like to thank the anonymous reviewers for their comments. This work was financially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Pierre Arbour Foundation.

## References

- [1] J. Piskorski, N. Stefanovitch, G. Da San Martino, and P. Nakov. “SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup”. In: *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. Toronto, July 2023, pp. 2343–2361.
- [2] J. Kiesel, M. Mestre, R. Shukla, E. Vincent, P. Adineh, D. Corney, B. Stein, and M. Potthast. “SemEval-2019 Task 4: Hyperpartisan News Detection”. In: *Proceedings of the 13th International Workshop on Semantic Evaluation (SemEval-2019)*. Minneapolis, Minnesota, USA, June 2019, pp. 829–839.
- [3] G. Da San Martino, A. Barrón-Cedeño, H. Wachsmuth, R. Petrov, and P. Nakov. “SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles”. In: *Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020)*. Barcelona (online), Dec. 2020, pp. 1377–1414.
- [4] D. Dimitrov, F. Alam, M. Hasanain, A. Hasnat, F. Silvestri, P. Nakov, and G. Da San Martino. “SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes”. In: *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*. Mexico City, June 2024, pp. 2009–2026.
- [5] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi. “Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking”. In: *Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*. Copenhagen, Sept. 2017, pp. 2931–2937.
- [6] K. S. R. Nayak and L. Kosseim. “Analyzing Persuasive Strategies in Meme Texts: A Fusion of Language Models with Paraphrase Enrichment”. In: *Proceedings of the 5th International Conference on Natural Language Processing and Applications (NLPA 2024)*. Sydney, Australia, 2024.
- [7] P. Goffredo, M. Chaves, S. Villata, and E. Cabrio. “Argument-based Detection and Classification of Fallacies in Political Debates”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*. Singapore, Dec. 2023, pp. 11101–11112.
- [8] J. Piskorski et al. *News Categorization, Framing and Persuasion Techniques: Annotation Guidelines*. Accessed 15-08-2024. 2023. URL: [https://knowledge4policy.ec.europa.eu/sites/default/files/JRC132862\\_technical\\_report\\_annotation\\_guidelines\\_final\\_with\\_affiliations\\_1.pdf](https://knowledge4policy.ec.europa.eu/sites/default/files/JRC132862_technical_report_annotation_guidelines_final_with_affiliations_1.pdf).
- [9] D. Dimitrov, B. Bin Ali, S. Shaar, F. Alam, F. Silvestri, H. Firooz, P. Nakov, and G. Da San Martino. “SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images”. In: *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*. Online, Aug. 2021, pp. 70–98.
- [10] W. C. Mann and S. A. Thompson. “Rhetorical Structure Theory: Toward a functional theory of text organization”. In: *Text - Interdisciplinary Journal for the Study of Discourse* 8.3 (1988), pp. 243–281.
- [11] E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber. “The Penn Discourse Treebank”. In: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon, Portugal, 2004, pp. 2237–2240.
- [12] R. Prasad, E. Miltsakaki, N. Dinesh, A. Lee, and A. Joshi. *The Penn Discourse Treebank 2.0 Annotation Manual*. Tech. rep. University of Pennsylvania, 2007. URL: <https://www.cis.upenn.edu/~elenimi/pdtb-manual.pdf>.
- [13] L. Carlson, D. Marcu, and M. E. Okurowski. *RST Discourse Treebank*. Linguistic Data Consortium. 2002. URL: <https://doi.org/10.35111/4w31-m996>.

- [14] R. Prasad, E. Miltsakaki, N. Dinesh, A. Lee, A. Joshi, and B. Webber. *The Penn Discourse Treebank 1.0 Annotation Manual*. Tech. rep. University of Pennsylvania, 2006. URL: <https://catalog.ldc.upenn.edu/docs/LDC2008T05/papers/pdtb-1.0-annotation-manual.pdf>.
- [15] R. Prasad, B. Webber, A. Lee, and A. Joshi. *PDTB 3 Annotation Manual*. Accessed: October 2024. Linguistic Data Consortium. 2019. URL: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
- [16] Y. Lei and R. Huang. “Discourse Structures Guided Fine-grained Propaganda Identification”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*. Dec. 2023, pp. 331–342. DOI: [10.18653/v1/2023.emnlp-main.23](https://doi.org/10.18653/v1/2023.emnlp-main.23). URL: <https://aclanthology.org/2023.emnlp-main.23/>.
- [17] A. Chernyavskiy, D. Ilvovsky, and P. Nakov. “Unleashing the Power of Discourse-Enhanced Transformers for Propaganda Detection”. In: *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024) (Volume 1: Long Papers)*. Malta, Mar. 2024, pp. 1452–1462. URL: <https://aclanthology.org/2024.eacl-long.87/>.
- [18] P. K. Choubey, A. Lee, R. Huang, and L. Wang. “Discourse as a Function of Event: Profiling Discourse Structure in News Articles around the Main Event”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*. Online, July 2020, pp. 5374–5386. DOI: [10.18653/v1/2020.acl-main.478](https://doi.org/10.18653/v1/2020.acl-main.478). URL: <https://aclanthology.org/2020.acl-main.478/>.
- [19] I. Rehbein. “On the role of discourse relations in persuasive texts”. In: *Proceedings of the 13th Linguistic Annotation Workshop*. Ed. by A. Friedrich, D. Zeyrek, and J. Hoek. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 144–154. DOI: [10.18653/v1/W19-4017](https://doi.org/10.18653/v1/W19-4017). URL: <https://aclanthology.org/W19-4017/>.
- [20] Z. Lin, H. T. Ng, and M. Kan. “A PDTB-Styled End-to-End Discourse Parser”. In: *Natural Language Engineering* 20 (2014), pp. 151–184.
- [21] J. Wang and M. Lan. “A Refined End-to-End Discourse Parser”. In: *Proceedings of the Nineteenth Conference on Computational Natural Language Learning (CoNLL-2015)*. Beijing, China, July 2015, pp. 17–24.
- [22] K. Thompson, A. Chaturvedi, J. Hunter, and N. Asher. “Llamipa: An Incremental Discourse Parser”. In: *Findings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*. Miami, Nov. 2024, pp. 6418–6430.
- [23] A. Maekawa, T. Hirao, H. Kamigaito, and M. Okumura. “Can we obtain significant success in RST discourse parsing by using Large Language Models?” In: *Proceedings of 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)*. Malta, Mar. 2024, pp. 2803–2815.
- [24] Google AI. *Gemini API Documentation: Prompting Strategies*. Tech. rep. Accessed: October, 2024. Google AI, 2023. URL: <https://ai.google.dev/gemini-api/docs/prompting-strategies>.
- [25] I. Rehbein, M. Scholman, and V. Demberg. “Annotating Discourse Relations in Spoken Language: A Comparison of the PDTB and CCR Frameworks”. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia, 2016.
- [26] N. F. Costa and L. Kosseim. *A Multi-Task and Multi-Label Classification Model for Implicit Discourse Relation Recognition*. 2024. arXiv: [2408.08971](https://arxiv.org/abs/2408.08971) [cs.CL]. URL: <https://arxiv.org/abs/2408.08971>.