Exploring the Role of Discourse Structure and Tropes in Detecting Enthymemes in Social Media Posts

Anonymous ACL submission

Abstract

This study investigates the linguistic characteristics signaling enthymemes—arguments with implicit premises or conclusions—in social media texts, focusing on their detection using computational methods. We address two primary research questions: (1) How effective are Rhetorical Structure Theory (RST) discourse features and online tropes in detecting enthymemes? (2) What micro-level rhetorical strategies characterize enthymemes in short texts? We augment a dataset of 313 tweets from the 2019 British electoral campaign, annotated for tropes and enthymeme presence, with automatically generated RST trees. Predictive models, including classical machine learning and transformer-based approaches (e.g., RoBERTa), are trained for enthymeme detection. Findings reveal that RST structural features, such as nucleus-satellite ratio, tree depth, and particular patterns of coherence relations, enhance model capacity to discern enthymeme presence. A rhetorical strategy involving JOINT, BACKGROUND, and minimal argumentative relations is identified as a key pattern in enthymeme encoding. While certain tropes (e.g., hidden_motives) correlate with implicit arguments, others are less reliable. Contributions include: (1) a novel dataset annotated for enthymeme presence, (2) an analysis of RST and trope feature efficacy, (3) a signal-based approach to discern enthymeme types, and (4) insights into micro-level discourse-driven rhetorical strategies for enthymeme detection.

1 Introduction

This paper explores the linguistic characteristics that signal the presence of arguments with missing premises or conclusions, also known as *enthymemes* (Walton, 2008; Feng and Hirst, 2011), in short texts from a computational linguistics perspective. The study is motivated by the need to be able detect enthymemes in social media, where they are a prevalent means of persuasion, as they can reduce the recognition of fallacies and false information, thereby facilitating deceptive argumentation and manipulation (Lombardi Vallauri et al., 2020). Below is an example of an enthymeme with a missing conclusion.

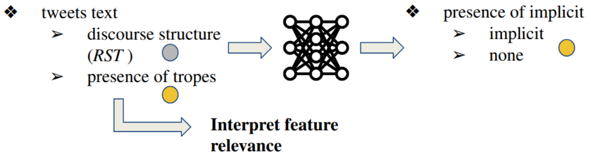

Figure 1: Overview of the workflow: gold circles represent gold annotations, while silver circles represent generated annotations that have been partially verified.

(1) “The UK Government missed their net migration target this year. The 37th year in a row they’ve missed their own target. Brexit was a movement for a British Britain, not a Britain packed to the rafters with immigrants who do not share our British values.” **[Implicit_distrust_experts_2149]**

- **Major Premise**: Brexit was intended to stop immigration.
- **Minor Premise**: The UK government continues to allow high levels of immigration.
- **Conclusion (implicit)**: Brexit has failed.

We frame the problem as a text classification task, where each text is classified as containing an implicit claim or conclusion, or not. Crucially, we explore different linguistic means of signalling the presence of enthymemes, focusing on the effectiveness of Rhetorical Structure Theory (RST) (Mann and Thompson, 1988) discourse characteristics and Online Tropes as feature types for training predictive models. While discourse structure captures the formal organization of textual content, online tropes reflect culturally situated, often implicit, rhetorical patterns typical of social media (Flaccavento et al., 2025).

For the purpose of our investigation, we enhance a dataset of tweets related to the 2019 British electoral campaign around Brexit, previously manually annotated with tropes by Flaccavento et al., (2025), by automatically generating RST trees for each tweet. We then manually annotate the data for the presence of enthymemes before training predictive models (classical machine learning and transformer based) for enthymeme detection. Figure 1 illustrates the main workflow adopted.

After reporting the model results, we conduct an error and success analysis to assess feature relevance. We begin by examining cases that have been correctly identified by models using discourse structure features but missed by those based on tropes. Next, we treat discourse and trope features as discriminative signals to compare their contributions to successful predictions. Finally, we reinforce our analysis by leveraging predictive models for error analysis using XGBoost (Shang et al., 2019).

Further, we discuss current views in the literature on implicit argument detection with respect to discourse features, where only the ambiguous role of discourse markers has been acknowledged. After showing the signalling relevance of discourse structure, in particular the nucleus to satellite ratio, the depth of certain relations, and their specific configurations, we argue that discourse features go beyond explicit discourse markers and discourse structure is indeed relevant to the task and warrants further investigation.

The role of tropes is then discussed in relation to recent models for enthymeme detection that have been enhanced with common sense knowledge. These models perform best when they rely on expectations about what typically occurs in certain stereotypical situations. We aim to investigate whether familiar tropes in a given context, function similarly to preconceived notions by guiding the inference of what is contextually relevant.

Following our results and analysis on signal types, we address how specific patterns of coherence relations, coupled with a minimal use of argumentative relations, contribute to the linguistic encoding of enthymemes and highlight a microlevel rhetorical strategy.

The main contributions of the paper are as follows:

- Publication of a dataset annotated for enthymeme presence.

- An analysis of RST discourse features and online tropes in relation to their contribution to model performance.

- An approach that treats these features as signals to investigate the types of enthymemes most discernible by predictive models.

- An investigation into the linguistic encoding of enthymemes from a discourse structure perspective.

2 Related Work

Enthymemes have drawn interest in argumentation studies because they can be formalized using propositional logic. Yet, this is difficult, as implied premises are often fallible or context-dependent (Walton, 2008; Razuvayevskaya and Teufel, 2017). Reconstructing them is also subjective, relying on background knowledge that varies across contexts, and becomes harder as arguments grow longer and implicit content increases (Sampson and Clark 2008; Becker, 2024).

While implicit premises and claims have been annotated in various ways, Sperber and Wilson, (2004) argue that readers infer them through relevance, using common sense or stereotypical knowledge (Walton, 2008). We explore how familiar tropes similarly guide inference.

Aside from Green, (2010) work, little research has examined how discourse structure signals implicit argumentation.

In computational approaches, implicit premises and claims have received some attention, particularly through recent generative methods for enthymeme reconstruction (Stahl et al., 2023; Chakrabarty et al., 2021). However, our focus here is narrower: we exclusively focus on the task of enthymeme detection.

Classical machine learning for detecting implicit premises often treats it as stance classification (e.g., explicit vs. implicit stance in reviews), using features like cue words and POS tags (Rajendran et al., 2016). Later work improved results with sentence embeddings (Schaefer and Stede, 2019).

Our approach differs by targeting diverse claims on nuanced topics like immigration, similar to Stahl et al., (2023), who detect implicit content presence in short texts beyond binary stances using a fine-tuned DeBERTA model.

Although common-sense-augmented models help detect implicit elements (Chakrabarty et al., 2021), it as been observed surface discourse markers can mislead systems (Becker, 2024). To our knowledge, no study has yet explored the role of discourse structure relations in automatically detecting implicit premises and claims in short texts.

3 Task Definition

Definition. An implicit argument can be defined as a chain of inferences in which one or more parts are left unstated. For example, a syllogism consists of three parts: (1) a major premise, (2) a minor premise, and (3) a conclusion. When one of these parts is missing, the syllogism is considered incomplete and is usually referred to as an enthymeme (Walton, 2008).

Problem definition. Given a short text, our goal is to assign one of two labels—Implicit or None—to indicate whether an implicit premise or conclusion is present.[1]

3.1 Annotation process

We annotate short texts, specifically tweets, which may contain one or more enthymemes, though typically only one. We do not aim to reconstruct the complete argument structure or provide the missing components. Instead, we focus on identifying whether a part of the logical syllogism, either a premise or a conclusion, has been left implicit.

Guidelines. Since almost any argument can be expressed as an enthymeme (Lippi and Torroni, 2015), we adopt a cautious approach to limit subjectivity and avoid redundancy (Becker, 2024). We annotate only cases where a single syllogistic component is missing, unless the major premise is common sense. If both major premise and conclusion are missing, we label it 'None', unless one of them is common sense. We require shared terms between premises to ensure minimal logical structure and exclude common-sense cases like "Rain gets you wet".[2]

The annotation involved three annotators and yielded a Cohen's Kappa inter-annotator agreement of 0.64 before consensus. The agreement was 0.54 when considering all three labels, including the distinction between premise and conclusion

Table 1: Gold distribution of None and Implicit cases in the annotated dataset.

Class	**N**	**%**
None	152	48.56
Implicit	161	51.44
-Conclusion	*84*	*26.84*
-Premise	*77*	*24.60*
All	313	100.00

4 Dataset

4.1 Tropes Dataset

In this study we make use of a subset of data derived from the Tropes dataset compiled by Flaccavento et al., (2025). Tropes have been defined in media studies: a storytelling device or shortcut that assumes the audience recognizes the situation (Gala et al., 2020). By "online trope" we mean such devices used in online discussions. These refer not to plots but to human situations, often implied rather than explicitly stated, yet clearly understood by readers.

The full Tropes dataset comprises 3,304 tweets annotated for 9 trope types. From this dataset, we extracted all tweets on the topic of Immigration including the trope annotations. To address class imbalance issues, we removed portions of tweets that contained no tropes, as these instances were overly dominant and would have skewed the training process. In all 313 tweets remained. The final distribution of tropes in the dataset used for our experiments is presented in Table 2.

4.2 RST Annotations

Rhetorical Structure Theory (RST; Mann and Thompson 1988) analyzes text coherence by segmenting it into elementary discourse units (EDUs), which are clause-like spans. EDUs are linked by coherence relations—such as ELABORATION, CONTRAST, CAUSAL, and TEMPORAL—forming a hierarchical tree that represents the text.

RST distinguishes two EDU types: "Nucleus", carrying essential content, and "Satellite" adding extra information. Some relations (e.g., JOINT, SAME-UNIT) are multi-nuclear.[3]

In our dataset, the annotations were further enriched with automatically generated RST annotations for each individual tweet. Each tweet was parsed using the DMRST discourse parser developed by Liu et al. (2021), which is based on

[1] We initially trained a model to distinguish between three labels: Implicit Premise, Implicit Conclusion, and None. However, due to severe class imbalance, the performance was too low to support a conclusive error analysis (F1 scores: 0.31 for premises, 0.34 for conclusions, and 0.64 for None with our best performing RoBERTa model).

[2] Full guidelines will be released with the dataset.

[3] In this article we use the harmonized set of 18 labels as described by Braud et al., (2017).

Table 2: Trope type distribution ranked by trope type frequency. In all 359 tropes are used: 270 out of all 313 cases in the dataset have a single trope, while there are 40 cases with 2 tropes, 3 with 3 tropes and 51 without a trope. The last two columns give the relative distributions across gold None and gold Implicit cases per trope type.

Rank	**trope type**	**N**	**% of all**	% None	% Impl.
1	defend_the_weak	78	21.73	47.44	52.56
2	wicked_fairness	68	18.94	44.12	55.88
3	hidden_motives	58	16.16	41.38	58.62
4	time_will_tell	33	9.19	39.39	60.61
5	distrust_experts	30	8.36	33.33	66.67
6	liberty_freedom	19	5.29	52.63	47.37
7	scapegoat	19	5.29	10.53	**89.47**
8	natural_trad._is_better	3	0.84	66.67	33.33
9	too_fast_too_early	0	0.00	0.00	0.00
0	no trope	51	14.21	**78.43**	21.57

XLM-RoBERTa-Base (Conneau et al., 2020). This multilingual, top-down system simultaneously performs EDU segmentation and RST tree analysis. The parser achieves state-of-the-art performance on span splitting, nuclearity determination, and relation classification, with accuracy scores of 88.2, 76.2, and 64.7 respectively. A model checkpoint is publicly available on their GitHub repository.

To enhance the RST annotations, we further applied a method introduced by Pastor et al. (2025, which fine-tunes the same DMRST parser using synthetic data to extract what they found to be a prevalent pattern of discourse relations found in social media: JOINT–JOINT–EVALUATION (JJE) sequences. Incorporating this improved parser is particularly relevant for our study, as it allows for a more precise analysis of discourse structures in social media. However, since the enhanced parser tends to over-predict this pattern—with a reported precision of 0.60 and an F1 score of 0.61—we manually reannotated the affected cases. This process led to the addition of 35 JJE instances that were not identified by the original DMRST parser.

5 Experiments

5.1 RST and Tropes Features

The RST discourse features are encoded in a 31-dimensional feature vector to capture Rhetorical Structure Theory (RST) discourse features of each tweet. The tropes are encoded using a one-hot vector of size 10, corresponding to the ten identified tropes.

Dimensions 1–18 represent 18 discourse relations (e.g., ATTRIBUTION, ELABORATION), each valued as the sum of depths where the relation appears in the RST tree (0 if absent).

Dimensions 19–30 encode structural features: total segments (19), non-span relations (20), average segment length in number of tokens (21), maximum tree depth (22), counts of Nucleus (23) and Satellite (24) spans, Nucleus-Satellite (25), Nucleus-Nucleus (26), and Satellite-Satellite (27) relations, leaf nodes (28), internal nodes (29), and the Nucleus-to-Satellite ratio (30, or 0 if no Satellite spans).

Dimension 31 indicates discourse marker presence (1 if any discourse marker from the list provided in Das and Taboada (2018) appears, 0 otherwise).

5.2 Training Models for Implicit Argument Detection

5.2.1 Classical Approaches

To predict the Implicit or None label, we employ both a Logistic Regression model and a Support Vector Machine (SVM) with a linear kernel, implemented via Scikit-learn (Pedregosa et al., 2011; Buitinck et al., 2013), using a feature ensemble of TF-IDF text representations, RST features, and weighted trope-specific features. TF-IDF vectorization (max_features=5000, stop_words='english') is used for preprocessing. Features are combined using sparse matrices for efficiency and evaluated with 5-fold cross-validation across 10 random states to provide sufficient data for later error analysis. Performance metrics (accuracy, precision, recall, F1-score) for both models are reported per fold and averaged in Table 3.

5.2.2 RoBERTa

Next we finetune a RoBERTa model (Liu et al., 2019) using PyTorch and Hugging Face's Transformers (Wolf et al., 2020), combining RoBERTa's [CLS] token embeddings, RST features, and trope features. Tweet texts are tokenized (max_length=128) to generate contextual embeddings, with the [CLS] token output capturing sentence-level semantics. RST and trope features are concatenated with the [CLS] output, passed through a linear classifier with dropout (0.1). The model is trained with 5-fold cross-validation over 10 random states, optimized using AdamW (lr=2e-5) (Loshchilov and Hutter, 2019) and with 4 epochs. Accuracy, precision, recall, and F1-score are reported in Table 3.

Table 3: Performance Metrics (Precision, Recall, and F1-score) for Logistic Regression, SVM, and RoBERTa models across feature sets (No features, RST, Tropes and RST + Tropes.

Model	Metric Type	No Features				RST				Tropes				RST + Tropes			
		Label	P	R	F1	Label	P	R	F1	Label	P	R	F1	Label	P	R	F1
Log. R.	Overall Avg	—	0.60	0.59	0.59	—	0.62	0.62	0.62	—	0.62	0.62	0.62	—	0.64	0.64	**0.64** (±0.02)
	Avg per Label	implicit	0.60	0.67	0.63	implicit	0.63	0.66	0.64	implicit	0.62	0.71	0.66	implicit	0.64	0.67	0.66
		none	0.60	0.52	0.55	none	0.62	0.59	0.60	none	0.63	0.53	0.58	none	0.64	0.61	0.62
SVM	Overall Avg	—	0.59	0.59	0.59	—	0.61	0.60	0.60	—	0.64	0.63	0.63	—	0.64	0.64	**0.64** (±0.01)
	Avg per Label	implicit	0.59	0.67	0.63	implicit	0.61	0.66	0.63	implicit	0.63	0.70	0.66	implicit	0.64	0.67	0.65
		none	0.60	0.52	0.55	none	0.60	0.55	0.57	none	0.64	0.56	0.60	none	0.63	0.60	0.62
RoBERTa	Overall Avg	—	0.67	0.67	**0.67** (±0.02)	—	0.66	0.66	0.65	—	0.66	0.66	0.65	—	0.65	0.64	0.64
	Avg per Label	implicit	0.68	0.70	0.68	implicit	0.66	0.67	0.67	implicit	0.66	0.71	0.68	implicit	0.66	0.65	0.65
		none	0.67	0.64	0.65	none	0.65	0.64	0.64	none	0.67	0.61	0.63	none	0.64	0.64	0.63

Table 4: Disagreement counts across RST and Tropes feature sets. In 109 cases (out of the total 313 cases) the use of RST features yields a different result from when Tropes features are used.

Class	N	RST None, Tropes Imp.	RST Imp., Tropes None
None	48	28	20
Implicit	61	36	25
-Conclusion	*30*	*22*	*8*
-Premise	*31*	*14*	*17*
All	109	64	45

Table 5: Comparison of RST Features for None vs. Implicit

Feature	None	Impl.	Diff
Gains (None > Implicit)			
nucleus_to_sat._ratio	1.14	0.92	+0.22
pres._manner-means	0.11	0.00	+0.11
pres._same-unit	0.07	0.00	+0.07
pres._textualorg.	0.07	0.00	+0.07
pres._attribution	0.07	0.00	+0.07
pres._enablement	0.11	0.04	+0.07
pres._comparison	0.04	0.00	+0.04
pres._contrast	0.07	0.04	+0.03
pres._explanation	0.11	0.08	+0.03
pres._elaboration	0.25	0.24	+0.01
Losses (Implicit > None)			
disc._marker_pres.	1.43	2.80	-1.37
num_relations	2.57	3.72	-1.15
num_segments	3.57	4.72	-1.15
nucleus_spans	2.21	3.36	-1.15
max_depth	2.82	3.96	-1.14
nucleus_nucleus_rel.	1.04	1.80	-0.76
avg_segment_length	12.18	12.84	-0.66
pres._joint	0.39	0.88	-0.49
nucleus_sat._rel.	1.54	1.92	-0.38
pres._background	0.32	0.64	-0.32
pres._evaluation	0.00	0.24	-0.24
satellite_spans	1.18	1.36	-0.18
pres._condition	0.00	0.04	-0.04

5.3 Results

Our results in Table 3 indicate that both classical machine learning models, Logistic Regression and SVM, benefit from the inclusion of additional features. Specifically, SVM performs comparably well when using either tropes or RST features individually, while Logistic Regression sees greater gains from tropes alone than from RST features. Notably, the best overall performance is achieved by RoBERTa without any added features. Incorporating external features in RoBERTa actually leads to a decline in performance, presumably because explicitly adding discourse features (e.g., mean-pooled representations) may introduce redundancy or interfere with the model's internal representations, thereby reducing its overall effectiveness.

6 Error/Success Analysis

6.1 Comparison of RST Features for None vs. Implicit

We further observe in Table 4 that the Logistic Regression models using only tropes or only RST features often make different predictions. Among the 109 cases where the two models disagree, RST performs better in identifying None instances. In contrast, the trope-based model is more effective in predicting implicit cases, particularly those involving implicit conclusions. Despite these differences, the overall number of correct predictions is quite similar between the two models, with RST correctly classifying 28 None cases and 25 Implicit cases, while the trope-based model correctly classifies 36 Implicit cases and 20 None cases. This balance is consistent with their identical F1 scores of 0.62.

Table 5 takes a closer look at the 53 cases correctly predicted by the Logistic Regression model enhanced with RST features, but not by the model using Trope features. The table separates these into 28 correctly classified None cases and 20 correctly classified Implicit cases, highlighting the differences in the corresponding RST features. This breakdown offers insights into which RST features were most effective in distinguishing between None and Implicit instances.

Some notable characteristics of the RST features used include the following:

1. The tweets containing implicit premises / claims demonstrate greater structural complexity, incorporating more discourse markers (2.80 vs. 1.43, difference: -1.37), segments (4.72 vs. 3.57, difference: -1.15), relations (3.72 vs. 2.57, difference: -1.15), nucleus spans (3.36 vs. 2.21, difference: -1.15), and deeper hierarchies (maximum depth: 3.96 vs. 2.82, difference: -1.14).

2. There is frequent use of relations such as JOINT (-0.49), BACKGROUND (-0.32) and EVALUATION, which contribute to a more subjective and critical attitude through layered arguments and suggestive language, as observed in example (2).

3. In contrast, None tweets exhibit a simpler and flatter structure, characterized by a higher nucleus-to-satellite ratio (1.14 vs. 0.92, difference: $+0.22$) and more direct relations such as ENABLEMENT ($+0.07$), TEXTUALORGANIZATION ($+0.07$), ATTRIBUTION ($+0.07$), and ELABORATION ($+0.01$), emphasizing clear and explicit messaging, as shown in example (3).

(2) I have zero tolerance for abuse of our immigration system. [BACKGROUND] Under my #NewPlanForImmigration, I want to ensure the British people have confidence in the system, [ELABORATION] including stopping those who threaten our national security [JOINT] & push dirty money around our cities.

(3) A National Crime Agency spokesman says [ATTRIBUTION] a 33-year-old Iranian national and a 24-year-old British man have been arrested in Manchester [CAUSE] on suspicion of arranging the illegal movement of migrants across the English Channel

We observe here that the RST feature-enhanced model makes its distinctions by making use of structural complexity (segments, nucleus-satellite order, depth) features, and specific relations like EVALUATION and BACKGROUND for Implicit tweets, versus simpler structures and direct relations for None tweets.

6.2 Interpreting Feature Relevance in RoBERTa's predictions

Here we zoom in on using features as signals for error or success analysis to better understand what the best performing system (RoBERTa, text-only) got right or wrong, rather than analyzing what feature proved to be helpful for predictions. The analysis given below is supported by the numbers in Table 6 in Appendix A. Though it is difficult to reason on the small dataset that we have and we should be careful with any generalizations, we do highlight the following characteristics which stand out when looking at the table:

High-Impact Framing: JOINT with BACKGROUND Relations. The combination of JOINT and BACKGROUND (where `presence_joint = 1` and `presence_background = 1`) stands out as a key feature in the analysis. This configuration is particularly dominant in Implicit arguments, with 40 instances and an F1 score of 0.89, ranking first among RST features for Implicit cases. It also appears in some None arguments, with two trope-based entries. For example, in the "time_will_tell" Implicit class, the model achieves a perfect F1 score of 1.0. The reason for this effectiveness lies in the way JOINT facilitates the listing of multiple claims, while BACKGROUND, similarly to EVALUATION, provides a context that suggests motives or critiques in Implicit arguments such as in example (4). This combination's flexibility, coupled with its relatively high frequency and performance across several tropes—such as "no_trope", "time_will_tell", and "hidden_motives"—makes it a particularly strong signal in the system's predictions.

(4) "OVER 700 Migrants today. [JOINT] YOU will have operations & tests cancelled [JOINT] YOU won't get a pension or elderly care [BACKGROUND] FACT Join @UKIP [EXPLANATION] because the Conservatives are killing us." **[Implicit_2805]**

Example (5) illustrates a typical use of an implicit premise, where three statements are presented without any explicit argumentative connective, serving as background to better frame the critique of the Conservatives. By doing so, the writer avoids directly stating the main premise, which is instead taken for granted, namely, that mass immigration is responsible for canceled operations and tests, as well as the unavailability of pensions and healthcare.

(5) "Innocent : [BACKGROUND] Seven-year-old Emily Jones was stabbed to death in a park in Bolton, [JOINT] and Albanian immigrant Eltiona Skana has been charged. [EVALUATION] Why isn't this story on the front page of every newspaper?" **[Implicit_2553]**

Similarly, we see here a complex structure that makes minimal use of argumentative relations. It connects multiple statements and frames them through a BACKGROUND relation, before concluding with a final point using the most frequent argumentative relation in these cases: EVALUATION. This relation is argumentative because it involves subjective interpretation or appraisal, often aimed at persuading or influencing the reader's perspective. In short, a common pattern is the assembly of non-argumentative relations (BACKGROUND and JOINT) leading up to a single argumentative relation. The reader then focuses on this final point and overlooks the deceptive logic of the preceding arrangement, which often rests on defeasible or debatable premises or claims, as in (4), where the media is portrayed as selectively reporting crime, potentially due to political bias or fear of fueling anti-immigration sentiment.

Tropes: Tropes are significant yet ambiguous signals in classification, with varying predictive reliability. The "scapegoat" trope strongly indicates implicit conclusions (5 to 7 instances, F1 scores 0.75–1.0, Table 6). Similarly, the "no_trope" label often signals the absence of an implicit argument, occurring in 5 None different feature types (6 to 9 instances, F1 scores 0.94–1.0), but this is not always the case, as high-impact framing discourse structures, such as max_depth = 2 and nucleus_spans = 2 (F1 = 1.0, None instances), provide a more accurate signal, achieving higher predictive precision than "no_trope" alone. In contrast, the "defend_the_weak" trope is less reliable, as it is distributed across 5 Implicit (7–14 instances, F1 0.66–1.0) and 5 None RST feature types (7–9 instances, F1 0.85–1.0), indicating that tweets with this trope are not consistently predicted, as its even spread reduces its discriminative power.

6.3 Predictive model for error analysis

Building predictive models for error analysis has recently become a popular practice for the explainability of deep learning models (Savinova and Hoek 2024; Liu et al. 2023). We use tropes and RST features to interpret which factors best predict RoBERTa's errors or successes in detecting implicit premises/claims. We train an XGBoost model on a training set combining RST features (e.g., discourse markers, tree depth) and trope features, with binary labels for RoBERTa's correct ('success') or incorrect ('error') predictions. Using the classification gain metric (Shang et al., 2019), we identify features most influencing RoBERTa's outcomes. The model achieves 0.65 accuracy.

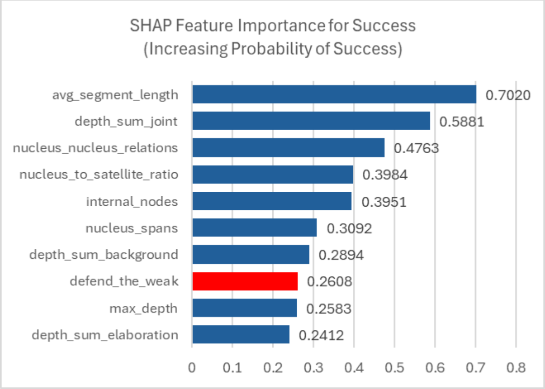

Figure 2: Feature relevance visualizations. SHAP values for XGBoost feature relevance for RoBERTa model. Tropes are color-coded in red.

What we observe from the graph in Figure 2 is that RST structural features are indeed picked up and contribute to the model's ability to predict successful classifications by RoBERTa. The most relevant features highlight aspects related to the position of JOINT relations (e.g., depth_sum_joint) and their prevalence. Specifically, high values for nucleus_nucleus_relations and nucleus_satellite_ratio suggest the presence of JOINT relations, which are by definition nucleus-nucleus. The relevance of these features in the figure thus suggests that larger JOINT patterns are likely to lead RoBERTa to successfully predict the presence of an implicit premise or claim.

In contrast, our manual inspection shows that tweets containing the "defend_the_weak" trope are not among the most successfully predicted. Its presence in the feature relevance graph (Figure 2 in Appendix A) suggests it contributes more to predicting RoBERTa's classification errors. This aligns with the earlier analysis indicating that certain tropes can be ambiguous, as they do not consistently signal the presence of an implicit premise.

Of note, the general absence of tropes among the most significant features in the graph in Figure 2 suggests that the presence of a trope only partly influences RoBERTa's preictions. This is reflected in RoBERTa's comparatively better performance in correctly identifying the absence of implicit premises or claims, unlike the trope enhanced SVM and Logistic Regression models, which tend to overpredict their presence when a trope is also present.

7 Discussion

We first report that incorporating the presence of tropes and RST characteristics as features improved results for classical machine learning approaches. More specifically, our feature and signal analysis suggests that it was structural characteristics such as the number of segments, the nucleus satellite ratio, the depth of certain relations, and their specific patterns that were most discernible by the RoBERTa model and the classical ML models. These discourse features go beyond simply annotating the presence of discourse markers and demonstrate that discourse structure has something meaningful to contribute to the task of enthymeme detection.

Beyond model performance, our results point to the existence of a particular micro-level rhetorical strategy, characterized by patterns of JOINT, BACKGROUND, and a minimal use of argumentative relations. To the best of our knowledge, no existing work in discourse analysis directly focuses on patterns of coherence relations in studying enthymemes. However, both Li and Xiao, (2021) and Pastor et al., (2024) note that texts containing multinuclear relations like JOINT and deeper structures are worth considering in the analysis of persuasion, as they find such elements to be frequent in their data. This aligns with the idea that when loosely connected statements are linked via the multinuclear JOINT, they can trigger an inference process in the reader, leading them to perceive the implicit content as their own (Reboul 2011; Lombardi Vallauri et al. 2020).

When it comes to tropes, just as Flaccavento et al. (2025), the authors of the dataset we use, have emphasized that tropes can exist independently of persuasion strategies, we observe that they can also exist independently of enthymemes. Our initial intuition was that tropes might function as perceived notions guiding inference, acting as common sense knowledge that could serve as a warrant for making an implicit claim, or even constituting the implicit premise themselves. Yet this is not always the case. Tropes may be more diffuse in context and cannot always be reduced to propositional elements of a syllogism. However, some tropes, such as hidden_motives and distrust_experts, can more readily be instantiated as premises or claims, for instance in the form "the media lies" or "the government is incompetent." These considerations lead us to suggest that certain tropes do foster a fertile ground for persuasion and implicit arguments, and should be more closely investigated in this specific context, while others may not.

8 Conclusion

In this paper we investigated the linguistic characteristics that may signal the presence of enthymemes. We augmented a dataset of tweets from the 2019 British electoral campaign on Brexit, previously annotated with tropes (Flaccavento et al., 2025), by automatically generating RST trees for each tweet. We then manually annotated the presence of enthymemes, resulting in a new dataset of 313 tweets labeled with implicit premises and claims. This dataset was then used to train predictive models, including classical machine learning and transformer-based approaches, for enthymeme detection.

We conducted an error and success analysis to evaluate feature relevance. Results show that classical machine learning models benefit from both trope and RST features. In particular, structural discourse features such as tree depth, nucleus-satellite ratio, and the positioning of certain relations contribute notably to performance. This supports the view that discourse structure is not limited to the explicit presence of discourse markers and deserves further investigation.

Our findings suggest a rhetorical strategy characterized by JOINT, BACKGROUND, and a limited use of argumentative relations. Although no previous work focuses on coherence patterns in enthymemes, our results align with studies such as Li and Xiao, (2021 and Pastor et al., (2024) which highlight the persuasive potential of multinuclear structures like JOINT

Lastly, some tropes foster implicit argumentation and merit closer study, while others are not consistently involved in implicit persuasion or enthymeme construction.

Limitations

A limitation of the research presented here is the size of the dataset, which provides limited insights into the types of patterns we observe, particularly with regard to model generalisation—even within the same type and topic distribution. This leads us to note that the phenomena we identify may be closely tied to our specific topic.

Moreover, it is important to note that we only work with generated data for RST annotations, and that better labeling of less recognized relations by parsers could have yielded different results.

Lastly, we reiterate that annotating enthymemes is a highly subjective task. Although we updated our guidelines to reflect a precise definition of enthymeme aligned with the logical syllogism, it would have been more robust to include more annotators from diverse backgrounds to further ensure annotation quality.

Acknowledgments

References

Maria Becker. 2024. *Building bridges. Reconstructing implicit information in argumentative texts using commonsense knowledge*. Number 61 in amades - Arbeiten und Materialien zur deutschen Sprache. Leibniz-Institut für Deutsche Sprache (IDS).

Chloé Braud, Maximin Coavoux, and Anders Søgaard. 2017. Cross-lingual RST discourse parsing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 292–304, Valencia, Spain. Association for Computational Linguistics.

Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake Vanderplas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. 2013. API design for machine learning software: experiences from the scikit-learn project. In *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases*, Prague, Czech Republic.

Tuhin Chakrabarty, Aadit Trivedi, and Smaranda Muresan. 2021. Implicit premise generation with discourse-aware commonsense knowledge models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6247–6252, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Debopam Das and Maite Taboada. 2018. Rst signalling corpus: a corpus of signals of coherence relations. *Language Resources and Evaluation*, 52.

Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 987–996, Portland, Oregon, USA. Association for Computational Linguistics.

Alessandra Flaccavento, Youri Peskine, Paolo Papotti, Riccardo Torlone, and Raphael Troncy. 2025. Automated detection of tropes in short texts. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 5936–5951, Abu Dhabi, UAE. Association for Computational Linguistics.

Dhruvil Gala, Mohammad Omar Khursheed, Hannah Lerner, Brendan O'Connor, and Mohit Iyyer. 2020. Analyzing gender bias within narrative tropes. In *Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 212–217, Online. Association for Computational Linguistics.

Nancy L. Green. 2010. Representation of argumentation in text with rhetorical structure theory. *Argumentation*, 24(2):181–196.

Jinfen Li and Lu Xiao. 2021. Neural-based RST parsing and analysis in persuasive discourse. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)*, pages 274–283, Online. Association for Computational Linguistics.

Marco Lippi and Paolo Torroni. 2015. Context-independent claim detection for argument mining. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, page 185–191. AAAI Press.

Yang Janet Liu, Tatsuya Aoyama, and Amir Zeldes. 2023. What's hard in English RST parsing? predictive models for error analysis. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 31–42, Prague, Czechia. 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. *Preprint*, arXiv:1907.11692.

Zhengyuan Liu, Ke Shi, and Nancy Chen. 2021. DMRST: A joint framework for document-level multilingual RST discourse segmentation and parsing. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 154–164, Punta Cana, Dominican Republic and Online. Association for Computational Linguistics.

Edoardo Lombardi Vallauri, Laura Baranzini, Doriana Cimmino, Federica Cominetti, Claudia Coppola, and Giorgia Mannaioli. 2020. Implicit argumentation and persuasion. *Journal of Argumentation in Context*, 9(1):95–123.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

Martial Pastor, Nelleke Oostdijk, and Martha Larson. 2024. The Contribution of Coherence Relations to Understanding Paratactic Forms of Communication in Social Media Comment Sections. In *JADT 2024 : 17th International Conference on Statistical Analysis of Textual Data*, Brussels (Belgium), Belgium.

Martial Pastor, Nelleke Oostdijk, Patricia Martin-Rodilla, and Javier Parapar. 2025. Enhancing discourse parsing for local structures from social media with LLM-generated data. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8739–8748, Abu Dhabi, UAE. Association for Computational Linguistics.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830.

Pavithra Rajendran, Danushka Bollegala, and Simon Parsons. 2016. Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 31–39, Berlin, Germany. Association for Computational Linguistics.

Olesya Razuvayevskaya and Simone Teufel. 2017. Finding enthymemes in real-world texts: A feasibility study. *Argument & Computation*, 8(2):113–129.

Anne Reboul. 2011. A relevance-theoretic account of the evolution of implicit communication. *Studies in Pragmatics*, 13(1):1–19.

Victor Sampson and Douglas B. Clark. 2008. Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3):447–472.

Elena Savinova and Jet Hoek. 2024. Subjectivity theory vs. speaker intuitions: Explaining the results of a subjectivity regressor trained on native speaker judgements. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 305–315, Bangkok, Thailand. Association for Computational Linguistics.

Robin Schaefer and Manfred Stede. 2019. Improving implicit stance classification in tweets using word and sentence embeddings. In *KI 2019: Advances in Artificial Intelligence*, pages 299–307, Cham. Springer International Publishing.

Erbo Shang, Xiaohua Liu, Hailong Wang, Yangfeng Rong, and Yuerong Liu. 2019. Research on the application of artificial intelligence and distributed parallel computing in archives classification. In *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 1267–1271.

Dan Sperber and Deirdre Wilson. 2004. Relevance theory. *Handbook of Pragmatics. Oxford: Blackwell*, pages 607–632.

Maja Stahl, Nick Düsterhus, Mei-Hua Chen, and Henning Wachsmuth. 2023. Mind the gap: Automated corpus creation for enthymeme detection and reconstruction in learner arguments. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4703–4717, Singapore. Association for Computational Linguistics.

D. Walton. 2008. The three bases for the enthymeme: A dialogical theory. *Journal of Applied Logic*, 6(3):361–379.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

A Appendix

Table 6: Final Ranking Table (Trope-Based)

Trope	RST Features	Number	F1	Rank
Implicit Instances				
defend_the_weak	nucleus_spans = 2, discourse_marker_presence = 1	14	0.666667	1
	presence_joint = 1, satellite_spans = 1	8	0.857143	2
	presence_joint = 1, nucleus_satellite_relations = 2	8	0.769231	3
	presence_evaluation = 1, presence_joint = 1	7	1.000000	4
	presence_joint = 1, num_segments = 5	7	0.923077	5
distrust_experts	presence_joint = 1, satellite_spans = 1	7	1.000000	1
	presence_background = 1, presence_joint = 1	6	1.000000	2
	presence_background = 1, discourse_marker_presence = 2	5	1.000000	3
	presence_background = 1, nucleus_satellite_relations = 3	5	1.000000	4
	presence_joint = 1, nucleus_spans = 3	5	1.000000	5
hidden_motives	presence_joint = 1, satellite_spans = 1	10	0.947368	1
	presence_joint = 1, discourse_marker_presence = 2	10	0.947368	2
	presence_background = 1, max_depth = 3, nucleus_nucleus_relations = 1	8	0.933333	3
	presence_background = 1, presence_elaboration = 1	7	1.000000	4
	nucleus_satellite_relations = 2, discourse_marker_presence = 1	7	0.923077	5
no_trope	presence_background = 1, presence_joint = 1	7	0.833333	1
scapegoat	presence_joint = 1, nucleus_spans = 4	7	1.000000	1
	presence_joint = 1, discourse_marker_presence = 3	6	0.909091	2
	num_segments = 5, discourse_marker_presence = 3	5	1.000000	3
	presence_evaluation = 1, satellite_spans = 1	5	0.750000	4
	presence_background = 1, presence_joint = 1	5	0.750000	5
time_will_tell	presence_background = 1, presence_joint = 1	7	1.000000	1
	presence_background = 1, discourse_marker_presence = 3	6	1.000000	2
	presence_background = 1, nucleus_spans = 3	6	0.909091	3
	presence_background = 1, max_depth = 3	6	0.909091	4
	presence_joint = 1, max_depth = 3	5	1.000000	5
wicked_fairness	presence_background = 1, satellite_spans = 1	9	0.800000	1
	max_depth = 3, nucleus_satellite_relations = 2	8	1.000000	2
	presence_background = 1, presence_joint = 1	8	0.933333	3
	presence_joint = 1, max_depth = 3, discourse_marker_presence = 2	7	1.000000	4
	nucleus_nucleus_relations = 1, discourse_marker_presence = 2	7	1.000000	5
None Instances				
defend_the_weak	presence_elaboration = 1, nucleus_spans = 2	9	1.000000	1
	presence_background = 1, satellite_spans = 2	9	0.875000	2
	presence_elaboration = 1, satellite_spans = 1	8	0.933333	3
	nucleus_nucleus_relations = 1, discourse_marker_presence = 1	8	0.857143	4
	presence_elaboration = 1, nucleus_satellite_relations = 3	7	0.923077	5
distrust_experts	presence_joint = 1, satellite_spans = 1	5	0.750000	1
	presence_joint = 1, nucleus_satellite_relations = 1	5	0.750000	2
hidden_motives	presence_joint = 1, nucleus_spans = 4	7	0.444444	1
	presence_background = 1, presence_elaboration = 1	6	0.666667	2
	presence_joint = 1, satellite_spans = 1	6	0.666667	3
	presence_joint = 1, discourse_marker_presence = 2	6	0.666667	4
	presence_background = 1, satellite_spans = 1	6	0.500000	5
liberty_freedom	presence_background = 1, presence_joint = 1	6	0.666667	1
no_trope	nucleus_spans = 2, discourse_marker_presence = 1	9	0.941176	1
	max_depth = 2, nucleus_spans = 2, satellite_spans = 1	7	1.000000	2
	max_depth = 2, nucleus_spans = 1, satellite_spans = 2	7	1.000000	3
	max_depth = 2, nucleus_satellite_relations = 1, nucleus_nucleus_relations = 1	6	1.000000	4
	max_depth = 1, nucleus_spans = 1, nucleus_satellite_relations = 1	6	1.000000	5
time_will_tell	presence_background = 1, presence_elaboration = 1	6	0.666667	1
	presence_elaboration = 1, num_relations = 5	6	0.666667	2
	presence_background = 1, presence_elaboration = 1, num_segments = 6	5	0.750000	3
	presence_background = 1, presence_elaboration = 1, num_relations = 5	5	0.750000	4
	nucleus_nucleus_relations = 2, discourse_marker_presence = 3	5	0.333333	5
wicked_fairness	presence_joint = 1, nucleus_spans = 3	8	0.933333	1
	presence_joint = 1, satellite_spans = 1	7	0.833333	2
	presence_elaboration = 1, satellite_spans = 1	6	1.000000	3
	presence_joint = 1, nucleus_satellite_relations = 1	6	0.909091	4
	nucleus_spans = 2, nucleus_nucleus_relations = 1	6	0.909091	5