



Dissecting Economic Op-Eds: An Annotation Schema for Editorials in Quality Newspapers

Javier Fernández-Cruz¹ 

Received: 23 December 2024 / Accepted: 5 May 2025 / Published online: 29 May 2025
© The Author(s) 2025

Abstract

This study examines the rhetorical structure and sentiment dynamics of 82 economic opinion editorials authored by prominent commentators. Using a novel annotation schema, we analyze Functional Discourse Units (FDUs), polarity, entities, and opinion holders to uncover the linguistic and rhetorical strategies employed in economic op-eds. The findings reveal a predominance of negative evaluations, highlighting the critical tone characteristic of the genre. Description/Analysis and Context FDUs are the most frequent, reflecting the op-eds' focus on situating economic issues within broader narratives. Annotation challenges, such as gradation of sentiment and irony, emphasize the complexity of evaluating economic discourse. Despite these challenges, high inter-annotator agreement in key categories supports the robustness of the schema. This study bridges computational linguistics and discourse analysis, offering valuable insights into economic journalism while laying a foundation for future research on evaluative language and sentiment analysis in specialized corpora.

Keywords Functional discourse units (FDUs) · Economic opinion editorials · Sentiment analysis · Rhetorical structure · Annotated corpora · Computational linguistics

Introduction

In the digital age, opinion editorials (op-eds) continue to wield significant influence, particularly in economics. Well-crafted opinion pieces can shift public opinion on policy issues, even among ideological opponents. Coppock et al. (2018) evidence that opinion editorials have the power to alter public opinion, even when read by a relatively small audience. This underscores the importance of communication channels like opinion editorials in shaping the decisions of companies and investors, who depend on analyzing consumer sentiment and economic indicators.

✉ Javier Fernández-Cruz
fernandezcruz@uma.es

¹ Facultad de Filosofía y Letras, Universidad de Málaga, Málaga, Spain

Given the volatility of economic discourse—i.e., the rapid shifts in economic narratives driven by market fluctuations, policy changes, and global events—accurately capturing textual opinions and attitudes remains pivotal for understanding public sentiment and influencing decision-making processes. Sentiment Analysis (SA), a widely used technique in Natural Language Processing (NLP), has become very influential in the financial sector. By extracting and quantifying sentiments from sources such as news articles, social media, and financial reports, SA aids in predicting market movements, guiding investment decisions, and enhancing risk management strategies (Oberländer et al., 2020). However, the unique features of the language of economics present challenges for SA. (Fernández-Cruz & Moreno-Ortiz 2023, 2024). While SA plays a key role in understanding public opinion, the complexity of economic discourse in opinion editorials demands a refined approach to capture the nuances of sentiment.

Opinion editors, as members of a discourse community (Swales, 1990), collaborate to shape public opinion through specialized language and distinct genres (e.g., editorials, comments, chronicles) (Melo & Assis, 2016). Within opinion editorials, the structure and function of economic discourse are particularly crucial to effective Sentiment Analysis (SA), as they encapsulate the complex interplay of language, tone, and context in shaping economic narratives. However, these features remain underexplored, highlighting a gap in research that needs to be addressed to fully leverage SA in economic contexts (Milutinović, 2021; Roslyng & Dindler, 2023; Zheng, 2014).

This study systematically investigates the rhetorical organization and lexical polarity of 82 opinion editorials, elucidating how discursive units and evaluative devices shape public opinion. Our annotation framework includes four layers: (a) polarity, (b) functional discourse units, (c) entities, and (d) opinion holders. Central to our framework is the concept of Functional Discourse Units from Egbert et al. (2021), which are discrete text segments that perform specific communicative functions (e.g., describe, evaluate, contextualize, etc.). Through the examination of how the structure of texts influences their reception and interpretation, we seek to enhance our understanding of the persuasive strategies employed by editors. Linguistically, this approach contributes to a deeper understanding of discourse structure, advancing research in sociolinguistics, discourse analysis, register variation, and computational linguistics. Furthermore, this study addresses a gap in the literature by proposing a computational model of annotation and offering a detailed exploration of textual organization in opinion editorials. Quantitative and qualitative insights into the textual organization of economic discourse have implications for public opinion formation, making this research also relevant for media studies, and economics.

Theoretical Framework

Textual Linguistics and Discourse Organization

Textual linguistics examines texts as holistic systems of communication, analyzing their syntactic coherence, contextual interplay, and functional completeness. In

economic op-eds, it provides a crucial lens for understanding how financial arguments are structured through rhetorical moves. Economic journalism relies not only on factual reporting but also on strategic linguistic choices that position arguments within broader ideological and policy debates. Integrating textual linguistics with discourse segmentation techniques enables a systematic analysis of how economic discourse is constructed and received. Beyond sentence-level structures, textual linguistics contextualizes texts within their social and communicative frameworks, evolving from early explorations of text grammars to more contextualized dynamic analyses of interactional discourse (Rocci, 2014).

The concept of “text”, rooted in the Latin *textus* (“tissue”), evokes notions of weaving and productivity. Understanding the role that texts play within their social and communicative contexts is essential. At the macro level, intertextuality refers to how texts are interconnected and shaped by other texts, drawing on cultural and linguistic influences. According to Kristeva (1986), “[e]very text is constructed as a mosaic of quotations” (p. 37), reinforcing the idea that texts are part of a larger network of meanings, influences, and references. At the micro level, text segmentation—whether through syntax, pragmatics, or genre-specific structures—reveals how individual parts of a text relate to different communicative functions, inviting readers to engage with multiple layers of meaning (Finlayson & Erjavec, 2017; Pak & The, 2018). Thematic progression, which tracks how themes evolve and develop throughout a text, further informs our understanding of how texts guide the reader’s attention and meaning-making process (Downing, 2001; Quasthoff et al. 2017). Textual segmentation and intertextuality together illuminate the interplay between textual structure, discourse function, and external references that shape interpretation. Examining segmentation highlights the boundaries and connections between discourse units, contributing to overall coherence and communicative intent.

As a basis for analyzing the text’s discourse organization, thematic progression, and rhetorical structures within news corpora (Fairclough, 2015; Halliday, 2014), we summarize the main theoretical approaches to textual segmentation, which offer distinct methodologies for identifying communicative goals and textual structures. In this study, we define *rhetorical strategies* as the linguistic and structural choices made by opinion editors in their texts to persuade, engage, or inform their audience. These strategies include explicit evaluative statements, metaphorical framings, intertextual references, and the strategic use of the citation of expert opinion. Understanding these rhetorical choices is essential in opinion journalism, where economic arguments must balance objectivity with persuasive intent.

The discourse structure of expository and argumentative texts has been studied using two primary corpus-based approaches: *bottom-up* and *top-down* (Vilar & Tolchinsky, 2022). *Bottom-up approaches* begin by automatically segmenting texts into discourse units based on linguistic features. These units are grouped by similarity to determine their communicative goals. Biber et al. (2004) focus on Vocabulary-Based Discourse Units (VBDUs), defined as topically coherent stretches of discourse identified on a linguistic basis. VBDUs are identified by the repeated use of specific vocabulary within a segment, and their communicative goals are subsequently described qualitatively. *Top-down approaches*, in contrast, start by defining discourse units based on the communicative goals of the genre. These units are then

analyzed linguistically. Upton & Cohen (2009) advocate this method as it provides clearer insights into the rhetorical structures of specific genres and their evolution throughout a text. In this approach, discourse units are identified by their purpose rather than linguistic features.

Toulmin's (1958) Argumentation Model consists of the *claim* (the writer's assertion) and the *grounds* (the evidence supporting the claim). Non-obligatory elements, such as *qualifiers*, *warrants*, *backing*, and *rebuttals*, provide further structure. This model offers a useful lens for understanding the persuasive structure of various discourse domains (e.g., legal, political, scientific, etc.), or in our case, economic op-eds, where claims about economic policies must be supported by expert evidence and rebuttals to counterarguments. For instance, in economic journalism, opinion editors often frame fiscal policies as either beneficial or harmful based on their perceived economic consequences, using Toulmin's framework of *claims*, *warrants*, and *backing* to construct compelling arguments, revealing the implicit assumptions that shape economic discourse.

Similarly, Swales' (1981) concept of *rhetorical moves* plays a pivotal role in genre analysis, originally in academic writing. These moves are distinct textual units designed to fulfill specific communicative functions within a text, thus guiding the discourse to meet the writer's objectives while adhering to the expectations of the discourse community. *Moves* are defined as a segment of text that carries out a particular function, such as (1) introducing a topic, (2) identifying a research gap, or (3) summarizing research within a context. Swales' framework has been adapted across diverse academic disciplines, such as scientific writing, tourism or pedagogy (Cenni, 2024; Revilla Vicente, 2008).

Finally, Mann et al.'s (1992) Rhetorical Structure Theory (RST) analyzes text organization by focusing on hierarchical relationships between elements like clauses and sentences. Key concepts include the distinction between the *nucleus* (central message, e.g., the main topic normally referred to in the news headline) and the *satellite* (supplementary information, e.g., supporting data or expert opinions). The interplay of *rhetorical relations*, such as *elaboration* (offering further arguments on the main point), *sequence* (presenting events in a specific order), and *motivation* (explaining reasons behind economic trends), helps convey the text's purpose, i.e., to inform, persuade, or critique. RST has applications in theoretical linguistics, discourse analysis, and natural language processing (NLP). Despite its influence, RST faces challenges in multi-level discourse analysis and capturing intentional structure in interactive dialogue systems.

Functional Discourse Analysis

The disciplines of Discourse Analysis (DA) (Jones, 2024), Conversation Analysis (CA) (Waring, 2018), and Genre Analysis (GA) (Bhatia, 2004, 2023) have significantly advanced the study of coherent discourse, each offering distinct perspectives. DA examines global discourse structures, such as narratives, focusing on their internal organization and linguistic markers (Labov & Waletzky, 1967; Schiffrin, 1987). CA investigates sequential processes within discourse, emphasizing interactional

dynamics and turn-taking mechanisms (Sacks, 1995). GA, informed by sociological and anthropological insights, explores genres as routinized responses to societal communicative challenges (Bauman, 2001).

Functional Discourse Analysis (FDA) is a discourse analysis approach grounded in Systemic Functional Linguistics (SFL), a framework developed by Halliday (2014). SFL views language as a social semiotic tool designed to meet communicative needs across diverse contexts. Central to this framework are three metafunctions: (1) the *ideational* metafunction, which represents experiences and constructs the external world; (2) the *interpersonal* metafunction, which governs social interactions and relationships; and (3) the *textual* metafunction, which organizes coherent messages within specific contexts. FDA shares similarities with DA, CA, and GA but also presents distinct differences. While DA focuses on global discourse structures and linguistic markers, CA emphasizes interactional dynamics and turn-taking mechanisms, and GA examines genres as responses to societal communicative challenges. FDA, grounded in SFL, analyzes how linguistic choices fulfill communicative functions within social contexts. This approach, while potentially aligning with CDA in uncovering the strategies behind power dynamics and ideologies, distinguishes itself by achieving specific communicative objectives within a given context.

Functional Discourse Units (FDUs) are coherent segments of discourse that perform distinct communicative functions within a text (Egbert et al., 2021). Unlike traditional sentence-based segmentation, FDUs provide a more flexible unit of analysis by capturing rhetorical intent rather than the mere annotation of syntactic structure. This distinction is particularly relevant in economic op-eds, where evaluation, analysis, and persuasion often extend beyond sentence boundaries. By identifying FDUs in economic discourse, we can systematically examine how authors construct arguments, reinforce viewpoints, and engage with readers—offering a more precise method for discourse analysis in economic journalism. Our adaptation tailors this model to economic opinion editorials by introducing several domain-specific modifications. In this context, a Discourse Unit (DU) is defined as: (a) coherent with an overarching communicative purpose, (b) marked by one or more specific communicative goals, and (c) self-contained, with a clear beginning and end.

The Genre of Op-Eds

The concept of genre is central to understanding communicative practices, both written and oral, and is extensively studied across fields like rhetoric, professional writing, and English for Specific Purposes (ESP). Genre studies examine the discursive and textual actions within communicative events alongside their forms and conventions. Approaches to genre studies are “used to describe texts within textual and social contexts, rejecting the idea that individual texts should be treated in isolation from either their or other texts” (Hyland, 2018, p. 507).

The term *op-ed* originally referred to columns positioned opposite the editorial page in newspapers. Today, it broadly refers to the genre of columns presenting the informed opinion of the writer on an issue relevant to a specific audience

(Orts Llopis, 2017). Op-eds are characterized by personal expression, allowing the writer to have complete control over content, while demanding clarity and engagement in presenting ideas. Previous research has highlighted cross-cultural and cross-linguistic variations in op-ed writing (Dafouz-Milne, 2008).

Economic op-eds, as a specific instantiation of the genre of op-ed writing, represent the intersection of individual expression and cultural influences. As Orts Llopis (2017) notes, these texts often blend specialized economic discourse with the writer's personal style. Key features of economic op-eds include:

- *Lexical-Semantic Choices*: Writers employ technical terms, idiomatic expressions, and neologisms to make complex economic concepts accessible and precise. A taxonomy from Orts Llopis (2015) on op-ed prose identifies the following lexical areas:
 1. *Technical or 'hard' words*: These terms belong to the theoretical field of economics and are highly abstract, used in specialized communication among economists. Examples include Latin rooted words such as 'recession' and 'bond', as well as culture-specific terms like 'stock' and "risk premium".
 2. *Neologisms*: Authors frequently coin new terms to make economic concepts more relatable, especially as economics becomes a universally relevant topic. These neologisms, whether created by the writer or derived from economic literature, bridge the gap between specialized and colloquial language. Op-ed writers coin expressions like "robber barons of yore" and "plutocrat-friendly policy", while terms like "mini-budget" capture specific economic phenomena. Additionally, neologisms from literature, such as eponyms like "Draghi moment" and phrases like "doom loop" and "fire sale" are increasingly adopted across languages.
 3. *Metaphorical Conceptualization*: Metaphors are used to frame economic crises, with choices reflecting individual and cultural influences. Many of these metaphors are conventionalized and no longer perceived as figurative, such as 'equilibrium', 'inflation', 'boom' and 'slump', which have become standardized terms in economic discourse.
 4. *Interpersonality and Metadiscourse*: The dialogic dimension of op-eds is reflected in how authors engage their readership. Through metadiscourse, writers position themselves within the broader discourse community, fostering a connection with readers and enhancing their argumentation.

These lexical strategies help make complex economic concepts more understandable and reflect broader cultural and communicative influences, enhancing the persuasive power of op-ed writing.

Additionally, effective persuasive writing in op-eds requires a careful use of facts, citations and opinions (Spencer, 2005) as writers seek to influence their audience through a careful combination of inducement and advocacy techniques. Both Orts Llopis (2017) and Dafouz-Milne (2008) suggest that effectiveness and

credibility are not always guaranteed, as the writer must establish a “credible textual persona or ethos” (Dafouz-Milne, 2008: 96) and engage in an effective dialogue with the audience. Our analysis of opinion holders in §5 will examine this interpersonal aspect of op-eds, quantifying the discursive and interpersonal layers. These insights emphasize the effectiveness of economic op-eds in communicating complex concepts to diverse audiences, demonstrating the genre’s adaptability and its significant role in shaping public discourse. Thus, by identifying FDUs in economic op-eds, we can systematically examine how authors construct arguments, reinforce viewpoints, and engage with readers.

Annotated Corpora

Corpus annotation involves systematically tagging and labeling linguistic features in texts (Sinclair, 2004), enabling the identification of syntactic structures, discourse markers, lexical choices, and stylistic devices in texts and genres, in general, and in news articles as the scope of this study, in particular. As a fundamental aspect of computational linguistics, annotation types—such as morphosyntactic tags, lemmas, syntactic parses, named entities, and sentiment (Kehoe & Gee, 2007)—provide crucial data for the development and testing of advanced language tools.

Additionally, computational text analysis utilizes annotated corpora to refine NLP algorithms (Finlayson & Erjavec, 2017) while contributing to linguistic theory by offering quantitative insights into language patterns. The integration of computational methods and linguistic annotation has thus become essential for exploring complex textual phenomena and advancing in the current age of Large Language Models (LLMs). Recent advancements, including the development of foundation LLMs like GPT (Brown et al., 2020), highlight the synergy between annotated corpora and NLP. These corpora enhance algorithm training and evaluation and deepen our understanding of language of different textual genres.

Related Work

The economic/financial domain presents specific challenges due to specialized language and a scarcity of domain-specific labeled data (Araci, 2019). Large pre-trained language models (or foundation models) have revolutionized NLP by serving as adaptable starting points for various tasks with minimal annotation, utilizing few-shot (i.e., using few labeled examples) or zero-shot learning (i.e., no direct training data) (Brown et al., 2020). These models have been further enhanced through pre-training on domain-specific data and self-supervised tasks tailored to target-specific tasks (Araci, 2019; Chalkidis et al., 2020; Gururangan et al., 2020).

Discourse markers, which manage discourse flow and structure, serve as learning signals for predicting implicit discourse and inference relations (Braud & Denis, 2016) and for learning general sentence representations that are transferable across various NLP tasks (Nie et al., 2019; Sileo et al., 2020). Prasad et al. (2017) explored discourse understanding by annotating discourse relations in the

Penn Discourse Treebank, focusing on causal, contrastive, and temporal relations. The SFU Opinion and Comments Corpus (Kolhatkar et al., 2020) exemplifies the annotation of online news comments.

The DART annotation scheme (Weisser, 2019) provides a comprehensive framework for annotating dialogue data with a focus on pragmatics. It employs XML tags to annotate dialogue at various linguistic levels, including syntax, pragmatics (such as speech acts), and semantico-pragmatics. The scheme categorizes syntactic units (e.g., declaratives, questions, imperatives, and discourse markers) and assigns attributes such as speech act type, polarity, mode of interaction, and topic.

Pascual (2023) analyzed a corpus of scientific dissemination, categorizing 27 pragmatic strategies into three macrocategories: *informative* (INF), emphasizing clarity and detail in communication; *promotional* (PRO), highlighting achievements and benefits to showcase the project's value; and *interactional* (INT), engaging diverse audiences through direct and indirect communication, utilizing digital affordances.

Cenni's (2024) study offers a genre analysis of 300 negative hotel reviews on TripAdvisor in French, Spanish, and Italian, revealing cross-linguistic uniformity in discursive strategies. The study finds that reviews across languages share similar structures, particularly in the use of functional *moves* like negative evaluations, positive remarks, background information, and recommendations. Intensification strategies, such as adverbs and metaphors, are frequently employed to emphasize negative experiences. Cenni also highlights the role of positive elements in negative reviews, supporting previous research and suggesting a standardized approach to online complaints across languages.

Building on the theoretical foundations discussed, the following section outlines the methodology for annotating economic op-eds.

Methodology

Research Questions

The main aim of the present research is to explore how economics op-eds structure their arguments and convey sentiment. Drawing on previous studies and their analytical tools (e.g., Functional Discourse Units analysis), we aim to test whether the rhetorical and linguistic features identified in earlier research are consistent in a specialized corpus of economic opinion pieces. To achieve this, we will address the following research questions:

1. What are the rhetorical and linguistic features of economic opinion editorials?
2. Which Functional Discourse Units (FDUs) constitute this specific textual genre, and what are the frequencies of the primary FDUs identified?
3. How are the entities and opinion holders represented in the text?

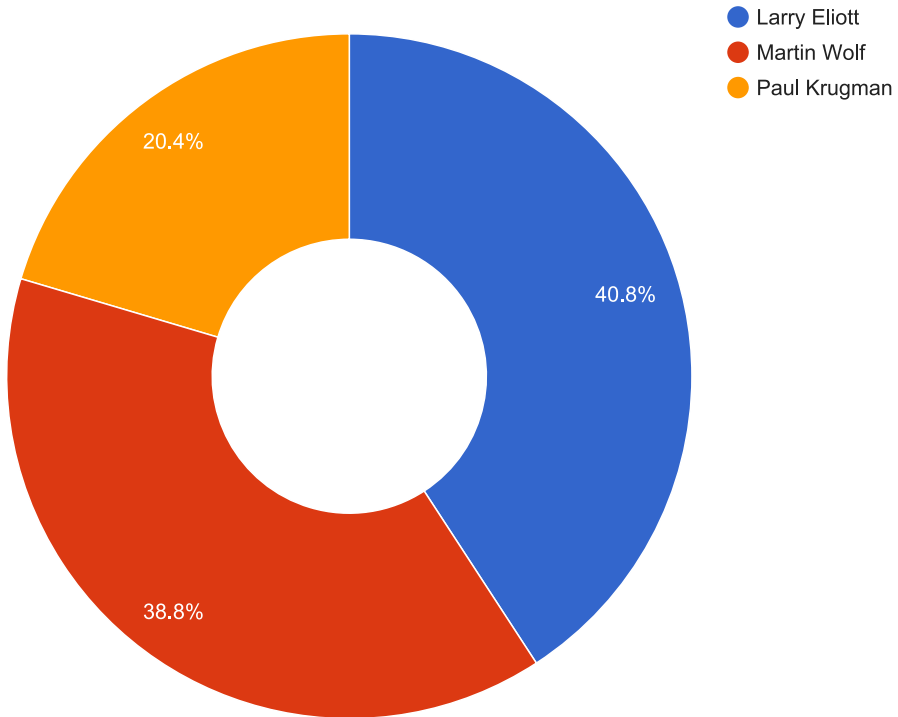


Fig 1 Distribution of token counts across the three authors in the Eco-DisparSA Corpus

Corpus Description: The Eco-DisparSA Corpus

We compiled the Eco-DisparSA Corpus¹, an annotated dataset comprising 82 opinion columns published between October 1st and December 31st, 2022. The corpus consists of 76,928 tokens, 67,071 words, and 3,400 sentences and is tagged by date and publisher (Fig 1). The final compilation follows the JSON Lines (JSONL) and XML format, ensuring ease of data handling, structured storage, and interoperability with computational analysis tools.

This corpus provides a focused examination of contemporary economic discourse as articulated by the chief opinion editors of their respective outlets: Larry Elliott (*The Guardian*, Keynesian), Martin Wolf (*The Financial Times*, centrist/pragmatic), and Paul Krugman (*The New York Times*, liberal/progressive in the American sense). Together, they represent a broad spectrum of economic perspectives. These authors are not only prominent columnists but also serve as key agenda-setters in economic journalism, determining the framing and prioritization of economic or financial issues in mainstream media. Their roles as senior economic commentators

¹ Eco-DisparSA is a compound acronym for *Economic Rhetorical-DIScursive PARsing and Sentiment Analysis*

ensure that their perspectives are widely disseminated and regarded as authoritative within both journalistic and policymaking circles.

EcoDisparSA was collected over a limited period, specifically from October 1st to December 31st, 2022, a timeframe deliberately chosen to capture economic discourse during a pivotal moment marked by concerns of inflation, energy crisis, and financial instability. These issues dominated global financial reporting during this period, making it a crucial window for analyzing the rhetorical and evaluative strategies employed in economic opinion writing.

The selected timeframe aligns with major economic events that shaped public discourse, ensuring the dataset captures the most pressing economic debates of the time. Due to this constrained sampling window, some authors contributed more articles than others, reflecting their publication frequency and editorial roles. This distribution reflects real-world media dynamics, where high-profile commentators publish more frequently than others.

As opinion editors, these authors set the tone and agenda for economic coverage in their newspapers, reinforcing the corpus's relevance as a model for high-profile economic commentary. Their regular contributions highlight their institutional role in defining economic narratives, making their work a representative sample of elite economic journalism rather than an arbitrary selection of op-eds. Our dataset, therefore, offers a representative snapshot of how economic journalism constructs and disseminates economic narratives, shaping public discourse in mainstream media.

The topics covered in Eco-DisparSA span prevalent themes in economic discourse. These include monetary and fiscal policies, the global energy crisis, inflation trends, and macroeconomic indicators such as bond yields and gross domestic product (GDP). The corpus also addresses historic political and economic events, including the short-lived tenure of British Prime Minister Liz Truss and its economic repercussions. Additionally, it examines broader global trends, incorporating perspectives on the economies of the United States, China, and the Russo-Ukrainian war.

Instruments

Prodigy (Montani & Honnibal, 2018; Montani et al., 2021), a Python-based annotation tool, played a central role in the annotation process. The tool was configured to use its incorporated spans.manual recipe that enables multi-layer annotation of text spans for any set of labels. The annotation workflow was meticulously streamlined and used a web-based interface, guaranteeing user experience, efficiency and reproducibility in the analytical process (Fig 2).

For the evaluation, we used Prodigy's integrated metric.iaa.binary recipe, which calculates Inter-Annotator Agreement (IAA) for binary classification tasks. Token-level annotations, particularly in span-based tasks, were assessed using pairwise F1 scores, which measure the precision and recall of overlapping annotations between annotators. The annotation tool also generates a confusion matrix, which breaks down true positives, false positives, and false negatives, providing insights into annotation consistency and systematic discrepancies.

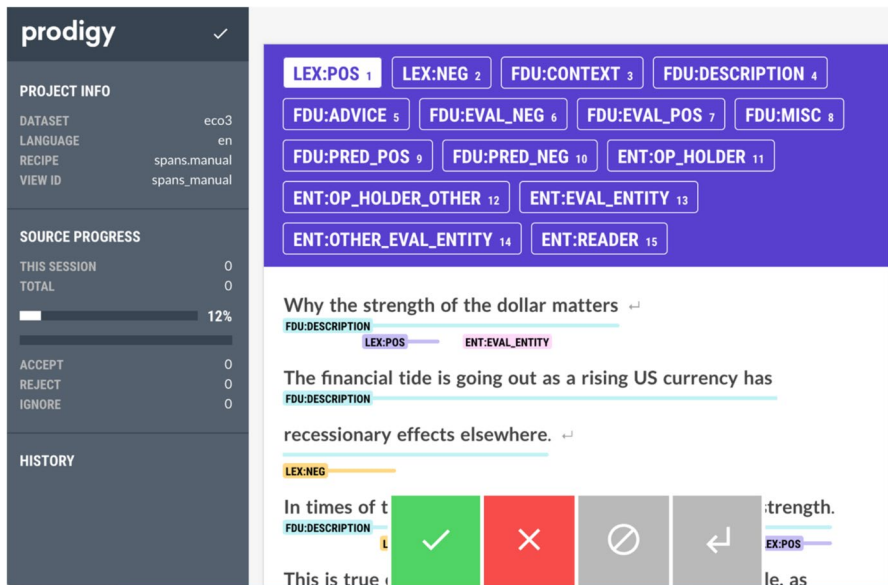


Fig 2 A general view of the Prodigy annotation interface

Procedure

The annotation process was conducted by the lead author, an expert in discourse analysis and sentiment annotation, in collaboration with a trained research assistant specializing in computational linguistics. Adherence to detailed guidelines was crucial to ensure accurate and meaningful results:

- **Proposition-Based Annotation:** Annotations were conducted on a proposition-by-proposition basis. Connectors such as ‘and’ or ‘nevertheless’ were generally excluded unless deemed essential for contextual comprehension. These connectors were only annotated when they were essential to maintain organizational clarity and accurately reflect distinct ideas or aspects.
- **Treatment of Punctuation:** Punctuation marks were generally excluded during annotation to minimize noise, as they rarely contribute to the rhetorical classification of FDUs. However, question marks and exclamation marks were retained due to their potential impact on sentiment and rhetorical intent. Question marks often signal uncertainty or skepticism, while exclamation marks convey emphasis or heightened sentiment. This approach aligns with standard NLP practice, where punctuation is selectively included based on its functional relevance to discourse analysis.

To ensure the consistency and reliability of the dataset, systematic measures were implemented:

- **Contextual Analysis:** Each proposition was analyzed within its surrounding context to understand its intended meaning and relevance. For example, if a context sentence referred to a prior event or action, this relationship was considered to maintain the integrity of the annotation. This ensured that each statement was interpreted accurately in its specific context rather than in isolation.
- **Focus on Key Entities:** Special attention was given to entities appearing in headlines or the first lines of text. These entities were often crucial for understanding the central theme of the text. For instance, in a news article headline, the primary entity (e.g., a political figure “Kwasi Kwarteng”, or organization “The Bank of England”) was annotated with particular care and their variations in the text were followed meticulously (“The Chancellor of the Exchequer” or “central bank”).
- **Consolidation of Related Segments:** Redundant annotation segments that address the same topic within a sentence were combined into a single segment, minimizing repetition while maintaining the integrity of key information. For example, the sentence, “Moreover, this has not been just against the currencies of emerging economies, but also against those of other high-income countries,” contains two propositions with the FDU:Analysis/Description. Therefore, we merged the two labels into one to streamline the annotation.

The manual revision process emphasized accuracy, consistency, and relevance, resulting in a dataset suitable for further analysis and insights. Given the complexity of FDUs, qualitative analysis was employed to verify annotation consistency and resolve disagreements through consensus. Discrepancies were resolved through discussion and alignment of coding criteria, reinforcing annotation consistency.

Following the annotation and revision of all texts, Prodigy generated a JSONL-format dataset containing the annotations in plain text. To facilitate analysis, a custom Python script processed the dataset, generating detailed visualizations and comprehensive statistical summaries.

Finally, to ensure dataset reliability and reproducibility in research, token-level annotations were evaluated using pairwise F1 scores to measure inter-annotator agreement (IAA) (see §4.6 for details) that provided insights into true positives, false positives, and false negatives in annotation spans.

Annotation Schema

The subsequent section provides a comprehensive discussion of the annotation labels used in each layer, which are justified based on the layers described by Liu (2015) for SA: *polarity*, *entity*, *opinion holder*, and *time*. The novelty lies in the inclusion of discourse aspects, specifically FDUs, which are also annotated.

Lexical Polarity (LEX)

This label identifies the sentiment polarity (positive or negative) of lexical units within the text. This annotation strictly focused on the lexical aspect, excluding considerations of pre-modifiers or contextual nuances. The approach adopted a

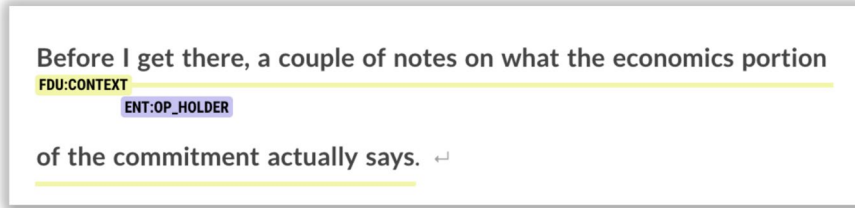


Fig 3 An example of CONTEXT

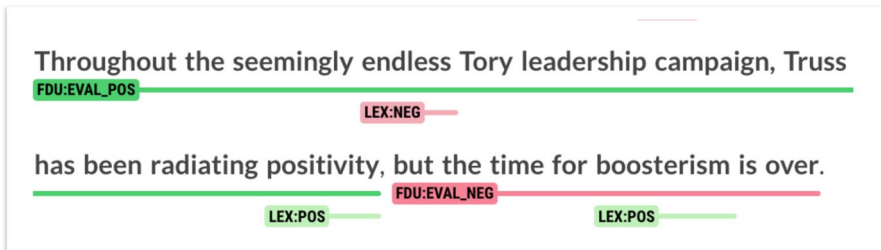


Fig 4 An example of EVAL_POS and EVAL_NEG in the same sentence

sentiment dictionary methodology to classify words based on their inherent positive or negative connotations. This systematic classification aims to provide a foundational understanding of sentiment orientation in the analyzed corpus.

Functional Discourse Units (FDU)

The FDU labels represent the communicative functions performed by different segments of text within the review (Figs 3 and 4):

- **ADVICE:** The author gives clear recommendations on economic management or urges readers to take specific actions.
- **CONTEXT:** Summarizes or refers to distant past events (economic or otherwise) from the author's viewpoint or introduces others' actions, opinions or statements.
- **DESCRIPTION/ANALYSIS:** Discusses or examines the recent or current state of economic affairs from the author's viewpoint, including hypothetical or conditional economic statements.
- **EVALUATION (EVAL_POS/EVAL_NEG):** Expresses a positive/negative opinion on the current state of economic affairs.
- **PREDICTION (PRED_POS/PRED_NEG):** Makes a clear negative/positive prediction about the future of the economy.

Opinion Holder and Entity (ENT)

The annotation framework includes Opinion Holders and Entities, two key layers essential for analyzing the evaluative structure of economic op-eds. Opinion holders represent the source of the evaluative stance, while entities identify the targets of evaluation within the text.

Opinion holders are classified into three categories. The primary opinion holder (OP_HOLDER) refers to the article's author, whose viewpoint shapes the editorial's stance. In contrast, other opinion holders (OP_HOLDER_OTHER) include external figures such as policymakers, economists, or referenced commentators, whose opinions contribute to the discursive interplay of authority and credibility. Finally, the reader (READER) label is used when the writer directly addresses or anticipates the audience's viewpoint, a strategy often employed for rhetorical persuasion.

The Entity layer identifies the subject of evaluation in economic discourse. The main evaluated entity (EVAL_ENTITY) represents the central focus of an op-ed, such as "the Bank of England" or "the Federal Reserve". Meanwhile, other evaluated entities (OTHER_EVAL_ENTITY) appear when the op-ed draws comparisons between different organizations, policies, or economic conditions to provide a broader contextual understanding.

Annotation Quality and Challenges: Evaluator Agreement and Coding Decisions

To ensure annotation reliability, IAA was assessed on a random subset of 60 annotated examples (representing 37.97% of the corpus' tokens). Table 1 summarizes IAA results. The pairwise F1 metric was used for consistency evaluation. High IAA was observed in categories like LEX:NEG (F1=0.96) and OP_HOLDER (F1=0.95), whereas FDU:PRED_POS (F1=0.34) exhibited lower agreement, indicating the challenge of annotating rare predictive statements.

Issues

The annotation of economic op-ed texts presented several intriguing challenges that enriched the analysis. One notable issue involved the presence of very short Functional Discourse Units (FDUs) containing two opposing evaluations. For example, phrases such as (1) present contrasting evaluations within a single, brief statement.

3. Glimmers of light [EVAL_POS] in a terrible year [EVAL_NEG] [EcoDisparSA-075]

The example demonstrates how economic op-eds often balance optimism with pessimism within a single evaluative statement. Here, the metaphorical "glimmers of light" introduces a positive assessment of economic prospects, while "a terrible year" simultaneously reinforces negative sentiment, reflecting the mixed rhetorical stance common in economic journalism. In these cases, annotators must carefully distinguish between the different evaluative components,

Table 1 Interannotator agreement (IAA) figures

Category	Pairwise F1	Support
LEX:NEG	0.96	1091
LEX:POS	0.93	436
FDU:ADVICE	0.49	27
FDU:CONTEXT	0.77	384
FDU:DESCRIPTION	0.80	648
FDU:EVAL_NEG	0.81	471
FDU:EVAL_POS	0.71	81
FDU:PRED_NEG	0.46	17
FDU:PRED_POS	0.34	5
ENT:EVAL_ENTITY	0.89	187
ENT:OTHER_EVAL_ENTITY	0.75	23
ENT:OP HOLDER	0.95	36
ENT:OP HOLDER_OTHER	0.87	100
ENT:READER	1.00	3

especially when embedded in a succinct expression. Such sentences often require careful contextual analysis to capture the nuanced shift in tone, making the accurate identification of sentiment both more challenging and crucial.

Another challenge was the gradation of sentiment, particularly when darker shades of sentiment were expressed. For example, statements like (2) reflect a deepening of negative sentiment, where the intensity of the evaluation is not simply binary but includes varying degrees of severity. These expressions introduce a level of complexity in SA, as they imply progression or escalation of negative outcomes. The presence of such gradations requires annotators to identify and account for the intensity of the sentiment, rather than simply marking it as negative. Future work should prioritize developing nuanced methodologies to systematically capture sentiment gradations, including variations in intensity and polarity shifts, helping to distinguish between different levels of negativity or positivity in economic discourse.

4. Worse may still be to come, [EVAL_NEG] possibly even much worse [EVAL_NEG] [EcoDisparSA-075]

Humor and irony were evidenced as complex due to their subtle nature and the ambiguity introduced by rhetorical devices. For instance, the use of sarcasm or irony in discussing policy issues created uncertainty about the author's true stance, often interpreted as a critical yet humorous commentary. Irony also emerged in discussions of economic predictions, where optimistic forecasts were tempered with acknowledged risks, reflecting a nuanced, critical perspective. Sarcasm, for example (3), is frequently used to critique policies while encouraging readers to engage in a more nuanced discussion. It relies on intertextual information as its meaning often hinges on the reader's knowledge of external references and shared cultural and ideological frameworks, or prior discourse, which provides the necessary backdrop for understanding the implied criticism:

5. [CONTEXT] What's my plan for the next two years? [EVAL_POS] I will be happy, and successful. [CONTEXT] What will I do to achieve these things? [EVAL_NEG] What are you, a Marxist? [EcoDisparSA-021]

While not exhaustive, these examples highlight the linguistic complexities inherent in op-eds and underscore the difficulty of discerning both the author's intent and the underlying sentiment. Analyzing these elements offered valuable insights into the rhetorical strategies used in op-eds and the inherently subjective nature of economic discourse.

Finally, the observed lower F1 scores in Positive Prediction (F1=0.34) and Advice (F1=0.49) reflect the inherent challenges of annotating nuanced and infrequent discourse features in economic opinion editorials. These discrepancies likely arise from subjectivity in interpretation, and the limited frequency of certain labels within the corpus. To address this, we propose refining annotation guidelines, enhancing annotator training with edge cases, and, in the future, leveraging automated tools to improve consistency. Despite these challenges, the high agreement in core categories underscores the overall robustness of the schema while highlighting areas for methodological refinement in future studies.

With a robust annotated dataset established, we proceed to analyze the distribution of lexical polarity, discourse functions, and evaluative structures within the op-ed corpus. The following section presents a quantitative assessment of these layers.

Data Analysis

Summary of Results

The analysis focuses on all four annotation layers: *Lexical polarity distribution*, *Functional Discourse Units (FDUs)*, *Entities* and *Opinion holders*. The findings provide insight into how economic op-eds employ textual and rhetorical strategies.

The analysis examined rhetorical organization and lexical polarity distribution in 82 economic opinion editorials. The spans per text in the corpus ranged from a minimum of 36 to a maximum of 236 characters ($\mu=121.28$, $Md=125.5$ and $Mo=143$). The standard deviation was 37.77, indicating moderate variability, while the variance was 1,426.52.

Description of Label Statistics

Lexical Polarity Distribution

To assess sentiment patterns, a count of positive or negative lexical items was conducted. As summarized in Fig 5, the corpus exhibits a strong inclination toward the use of negative words (2,985 occurrences, 70.99%) over the use of positive polarity tokens (1,120 occurrences, 29.01%).

The ratio of negative to positive polarity (2.66:1) corroborates the predominance of critical perspectives in economic op-eds. This aligns with the high frequency of Negative Evaluation and Negative Prediction FDUs, reflecting the op-eds genre focus on highlighting risks, challenges, and unfavorable scenarios.

Negativity is a prominent news value, in agreement with Bednarek and Caple (2017), with negative events like crises and conflicts often prioritized by media outlets. This inclination toward negativity likely stems from psychological factors, such as the urgency of adverse events, coupled with journalistic conventions that prioritize crisis-driven narratives. In addition, journalists may favor negative stories due to their perceived importance or alignment with prevailing narratives.

Functional Discourse Units (FDUs)

The corpus was annotated for FDUs based on our adaptation of Egbert et al. (2021), defining FDUs as contiguous text segments fulfilling distinct communicative purposes. Seven primary FDU categories were identified: Advice (196 occurrences), where authors propose recommendations or solutions; Context (973 occurrences), providing historical, or background information; Description/Analysis (1,492 occurrences), detailing factual accounts or detailing economic phenomena.

Across all 4,834 FDUs (an average of 58.85 per op-ed), economic op-eds prioritize critical engagement, with Negative Evaluation (1,552 occurrences, 32.11%) being the most dominant category. These evaluations critique policies, individuals, or events, reinforcing the genre's role in holding institutions accountable and addressing public concerns during economic uncertainty. The prevalence of Negative Prediction (170 occurrences, 3.52%) over Positive Prediction (75 occurrences, 1.55%) further underscores the risk-oriented stance of financial journalism, as op-eds are more likely to anticipate negative economic outcomes than optimistic ones. Figure 6 highlights that Negative Evaluation (32.11%) and Description/Analysis (30.86%) dominate economic op-eds, emphasizing critical assessment of policies and detailed exposition of financial conditions. Context (20.13%) provides historical and economic framing, while lower frequencies of Positive Evaluation (7.78%), Advice (4.05%), and Positive Prediction (1.55%) may indicate that financial journalism focuses more on risk assessment than prescriptive recommendations. This rhetorical strategy prioritizes problem identification, critique, and economic risk analysis, shaping public discourse on financial policies and trends.

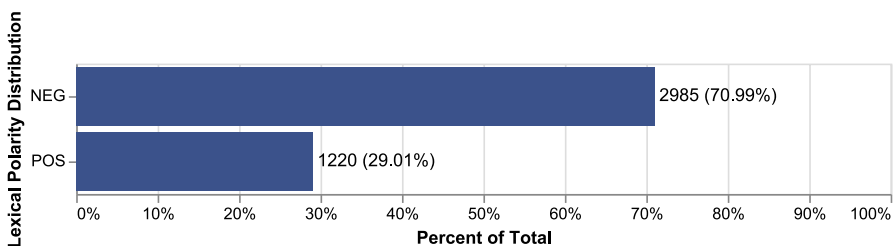


Fig 5 Frequency distribution of positive and negative lexical items within the Eco-DisparSA Corpus

Examples 4–8 illustrate how economic op-eds construct negative evaluations (EVAL_NEG), describe economic trends (DESCRIPTION/ANALYSIS), provide historical and political background (CONTEXT), and, less frequently, offer direct policy recommendations (ADVICE).

4. [EVAL_NEG] The Bank was concerned that it threatened the financial health of Britain’s biggest pensions and insurance companies, [DESCRIPTION/ANALYSIS] which together manage trillions of pounds of people’s cash. [EcoDisparSA-004]

5. Moreover, [EVAL_NEG] trade is just not growing as fast as before. [DESCRIPTION/ANALYSIS] That is in part a natural development. [DESCRIPTION/ANALYSIS] Globalisation could not grow as fast as before. [EVAL_POS] But it is still at work. [EVAL_POS] The world economy also continues to grow. [EVAL_POS] Our ancestors would find this extraordinary. [EcoDisparSA-075]

6. [DESCRIPTION/ANALYSIS] Services are a very heterogeneous group of activities, some of which require movement of people (tourism, for example). But activities in the exceptionally dynamic category of “other commercial services” (OCS) can, in large part, be supplied virtually. [EcoDisparSA-025]

7. [CONTEXT] Liz Truss was facing calls from jittery Conservative MPs on Wednesday to sack Kwarteng or face a mutiny after the Bank of England’s emergency intervention prompted comparisons with Black Wednesday, [CONTEXT] the day in September 1992 when John Major’s Tory government was humbled by speculators led by George Soros. [EcoDisparSA-004]

8. [ADVICE] The government should also subsidise the price of gas imports to these controlled levels. [ADVICE] These controls (and subsidies) should end when prices of imports fall back, [PRED_POS] as they surely will. [EcoDisparSA-052]

Headlines summarize the author’s main statement, providing essential information in an engaging manner. They are designed to capture the reader’s attention while offering a concise overview of the topic. The analysis of the FDUs in the op-ed’s headlines (Fig 7) reveals a clear emphasis on descriptive and evaluative content.

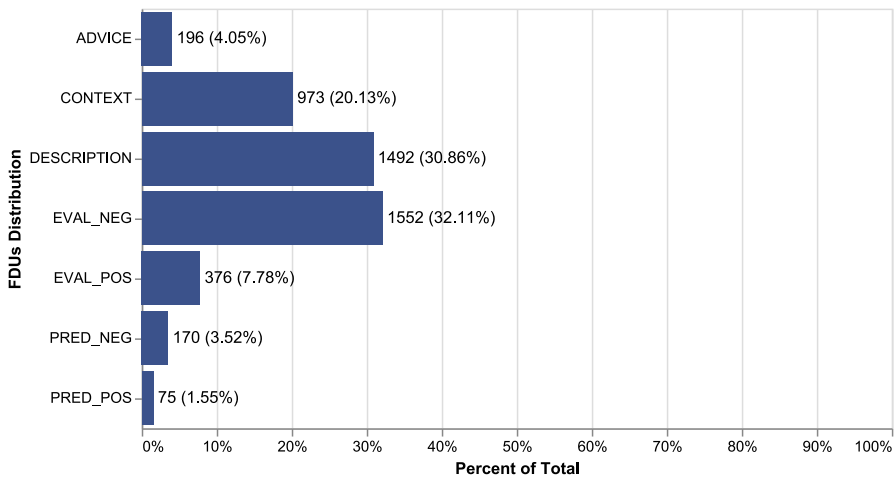


Fig 6 Relative frequency of Functional Discourse Units (FDUs) across the annotated corpus

The most frequent category, Description/Analysis (36 occurrences), indicates that headlines primarily convey factual and explanatory information about reported events. This aligns with the core function of news headlines—to inform readers quickly and clearly. Evaluation (28 occurrences) follows, highlighting how headlines often include critical assessments of events, individuals, or policies, reinforcing the role of shaping public opinion. Context (8 occurrences) may imply that while some headlines provide situational or background information, their main function remains immediate reporting of events rather than contextualization. The aggregated Prediction category (6 occurrences) suggests that the headlines rarely speculate on future outcomes, focusing on reporting and analysis rather than forecasting. Finally, Advice (4 occurrences) is the least common, reflecting the journals' preference for informing and evaluating over directly prescribing solutions. Some examples (9-11) follow:

9. [EVAL_NEG] UK economic growth remains tepid despite sizzling temperatures [EcoDisparisa-07]

10. [DESCRIPTION] Why the strength of the greenback matters [EcoDisparisa-019]

11. [PRED_NEG] Bank of England's wait-and-see approach unlikely to calm nerves [EcoDisparisa-03]

The analysis of the FDUs in the final lines of op-eds (Fig 8) reveals a distinct rhetorical strategy aimed at reinforcing critical perspectives and summarizing key insights. The prominence of Evaluation in concluding remarks (35 occurrences, 42.68%) underscores the genre's tendency to frame economic discourse through critique, influencing how readers interpret the column's main message.

Description/Analysis (18 occurrences, 21.95%) frequently appears in conclusions, serving to summarize key facts or clarify complex arguments. Interestingly, Advice (13 occurrences, 15.85%) occurs more frequently in final lines than in headlines, suggesting a subtle shift toward offering recommendations once the putative reader has engaged after detailed analysis. Predictions—both negative and positive (10 occurrences, 12.12%)—also emerge in closing statements, contributing to the persuasive function of op-eds by suggesting likely economic trajectories. Overall,

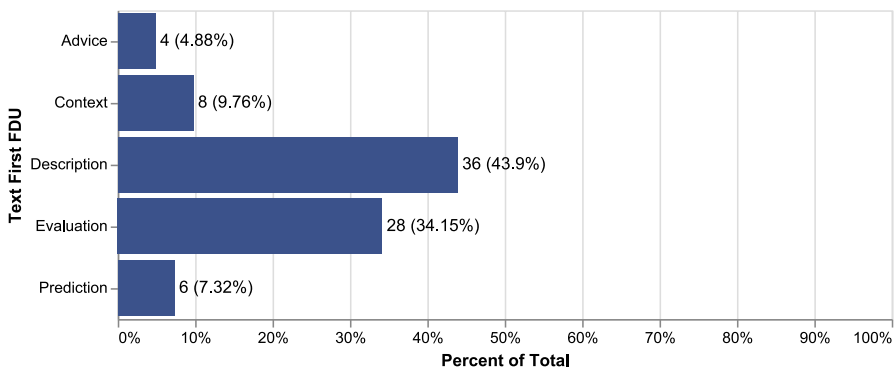


Fig 7 Breakdown of Functional Discourse Units (FDUs) in the headlines of economic op-eds

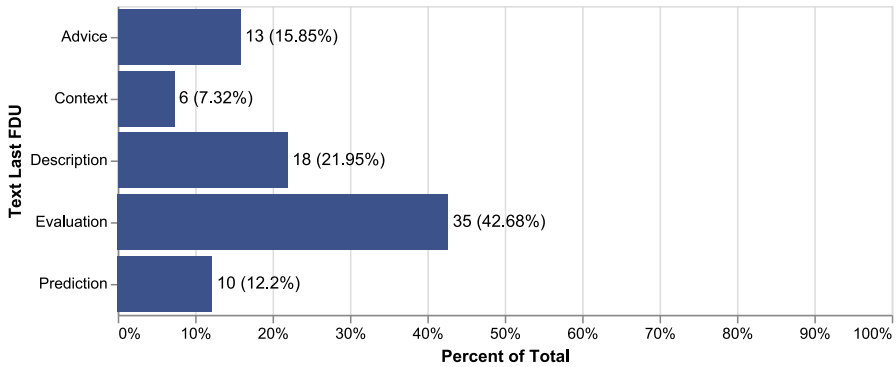


Fig 8 Breakdown of Functional Discourse Units (FDUs) in the last line of each document

the FDU distribution in final lines contributes to the persuasive strategy that reflects a tendency toward reinforcing critical viewpoints and, in multiple cases, making forward-looking statements. Illustrative instances (12–14) are listed next:

12. [DESCRIPTION/ANALYSIS] Forcing Asian countries to choose between China and the west would be similarly counterproductive. [DESCRIPTION/ANALYSIS] Again, who are our “friends”? Friendships change. [CONTEXT] Not so long ago, Vietnam was an enemy. [EcoDisparSA-041]

13. [ADVICE] It is not late, though, for Truss (or Sunak) to ensure it is at the milder end of the spectrum. [EcoDisparSA-065]

14. [CONTEXT] George Buckley, economist at Japanese bank Nomura said: [PREDICTION] “Florists in London will do well and sales of coffee will go through the roof. [PREDICTION] But in aggregate people aren’t going to be producing as much.” [EcoDisparSA-071]

Entities and Opinion holders

The final layer of analysis in the annotation schema is the Entity layer (Fig 9), which includes mentions of evaluated entities (EVAL_ENTITY) or comparable entities of the same category (EVAL_ENTITY_OTHER). For simplicity, the last two layers are grouped in Fig 9. Specifically, there are 577 mentions of the evaluated entities—the primary entity being evaluated in each news item—and 125 mentions of comparable entities, which serve as points of comparison.

This indicates a strong focus on direct evaluations of specific entities, with a notably lower tendency to contrast them with comparable entities. The prominence of EVAL_ENTITY suggests that the articles emphasize assessing key figures, organizations, or policies central to the narrative. In contrast, the relatively lower frequency of EVAL_ENTITY_OTHER implies that comparisons are used more selectively, likely to highlight distinctions or provide additional context. Thus, we can infer that the rhetorical strategy centers on direct evaluation while only occasionally leveraging comparative analysis. Example 15 illustrates both labels appearing in the same sentence:

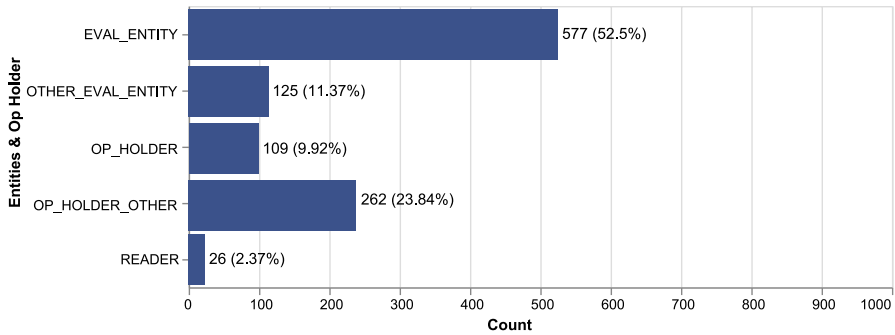


Fig 9 Counts for the layers of Entity and Opinion holder

15. The [EVAL_ENT] pound fell sharply across the world's currency markets, dropping below the \$1.09 level against [EVAL_ENT_OTHER] the dollar.

The annotation schema identifies explicit references to opinion holders, with 109 mentions of OP HOLDER, referring to the primary opinion holder, typically the author or the main source of the opinion. In contrast, there are 262 references to OP HOLDER_OTHER, which denote the opinions of other individuals or groups, such as experts, institutions, or public figures. This 2.2:1 ratio reveals that the articles place greater emphasis on the perspectives of external sources rather than the author's own viewpoint. By incorporating multiple voices, the author strengthens the credibility of the narrative and enriches the overall discourse, presenting a broader and more balanced view of the issues discussed.

The substantial number of references to other opinion holders suggests that the articles rely heavily on external opinions (generally other experts') to justify the author's argument, positioning these voices as authoritative and central to the article's persuasive power. This rhetorical strategy provides a more nuanced approach to presenting complex issues, allowing the text to appeal to a diverse audience and incorporate expert insights which enhances its objectivity and depth.

In addition to these external references, 26 mentions of the reader (0.31 per document) indicate direct engagement with the audience. These references often anticipate or address the reader's perspective, creating an engaging space within the text where the reader is invited to reflect on the content. This direct engagement is scarce but crucial for the persuasive nature of opinion-based articles, as it encourages active participation and critical thinking on the part of the reader. Some examples (16-18) follow:

16. According to the [OP HOLDER_OTHER] Treasury view of the world, Truss's go-for-growth approach represents a colossal gamble... [Eco-DisparSA-082]

17. Obviously, [OP HOLDER] I don't believe their assurances. More important, neither do [OP HOLDER_OTHER] financial markets. [Eco-DisparSA-030]

18. When [READER] you're telling stories about the economy, housing almost always looms large.

Conclusion

This study provides a novel contribution to economic discourse analysis by systematically annotating sentiment lexicon, FDUs and entities in op-eds. Through the development of a multi-layer annotation framework, it offers a computationally grounded and discourse-driven approach for analyzing how economic arguments are structured, framed, and evaluated. These findings offer better understanding of persuasive economic discourse while offering a linguistic resource that can improve Natural Language Processing (NLP) applications for sentiment and argumentation analysis.

The Eco-DisparSA corpus, as an electronic linguistic resource, serves multiple research objectives. Its annotated format allows for detailed SA at the textual level, facilitating the study of language strategies used to frame economic narratives. The structured representation of tokens, sentences, and documents enables a systematic comparison of authorial styles, thematic foci, and discursive strategies. The corpus is particularly suited for examining how economic journalism constructs meaning and conveys ideological or analytical stances within public discourse. Beyond computational linguistics, the corpus is a valuable resource for researchers in journalism and economics.

By offering a comprehensive, multi-author perspective on economic debates during a critical period, this dataset opens new avenues for exploring how opinion journalism shapes and reflects public understanding of complex economic phenomena. While it captures the rhetorical and evaluative strategies of influential elite economic opinion journalism, its composition may reflect broader gender and geographic disparities in quality press, signaling shifts in economic commentary. Major newspapers, however, remain largely dominated by male voices, particularly in editorial leadership. Semi-automatic annotation, using this annotation framework and dataset, could expand the scope by incorporating a wider range of ideologies, female economic commentators, and opinion writers from non-Anglophone markets, offering a comparative perspective on how gender and cultural factors shape economic discourse. This expansion would not only enhance generalizability but also contribute to a more inclusive understanding of economic journalism.

The analysis of 82 texts reveals a strong prevalence of negative lexical polarity (70.99%) in economic op-eds, highlighting the critical tone of economic discourse. This suggests that economic journalism serves not only as a source of information but also as a rhetorical tool that shapes public perceptions of risk and influences economic sentiment. Functional Discourse Units (FDUs) are mostly evaluative (32.11%) and contextual (20.13%), effectively positioning arguments within a broader economic narrative. The significant presence of third-party opinion holders, with a 2.2:1 ratio, indicates that authors strategically use expert attribution to enhance the credibility of their arguments and persuade their readership. Overall, for policymakers and media analysts, understanding these sentiment trends is essential in evaluating how the interplay of multiple voices, facts, and analyses shapes public opinion and investor confidence in economic matters.

Likewise, computational linguists can leverage these insights to enhance domain-specific sentiment analysis models, improving financial risk forecasting and automated economic discourse analysis. To fully unlock the potential of function-to-form analysis, advanced computational tools are needed to automatically detect pragma-discursive features—such as evaluative stance markers (e.g., “significant improvement”, “pressing need”) and discourse relations (e.g., contrastive markers like “however” or causal connectors like “therefore”)—with high precision. Developing such automated annotation models for FDUs and evaluative markers using large language models (LLMs) could greatly enhance the scalability and reproducibility of sentiment analysis (SA) research, as certain rhetorical elements may carry greater sentiment weight depending on their discourse function. Moreover, integrating discourse-aware models into financial NLP applications could improve sentiment-based market forecasting and media influence assessments, benefiting both researchers and industry professionals.

Future research could expand its scope in several key areas. First, expanding the corpus of economic op-eds to include a broader range of topics (e.g., op-eds on the stocks market) and perspectives would strengthen the generalizability of the findings. Particularly, incorporating diachronic analyses would provide insights into how economic sentiment evolves over time, particularly in periods of economic crisis vs. growth. Second, a comparative study across different languages could reveal cultural differences in economic discourse, offering a global perspective on discourse framing. In fact, one limitation of this study was its focus on the most influential English-language op-eds, which may not capture the richness of cross-cultural variations in rhetorical strategies. Lastly, integrating Large Language Models (LLMs) for automated annotation could improve the scalability and reproducibility of discourse segmentation tasks in computational linguistics. These directions underscore the interdisciplinary relevance of the study, with applications in media studies, sentiment analysis, and economic communication.

Acknowledgements The author gratefully acknowledges Antonio Moreno Ortiz for his invaluable technical support, Irina Muñoz Toala for her assistance with the annotation and evaluation during the initial stage of this paper, and Carla Fernández Melendres for her insightful feedback during the revision process. The author also thanks the anonymous reviewers for their valuable feedback and constructive suggestions during the review process. This research has been funded under the Margarita Salas research program from the Spanish Ministry of Universities, funded by the European Union-Next Generation EU. <https://www.uma.es/servicio-de-investigacion/info/129684/ayudas-para-la-recualificacion-del-sistema-universitario-espanol/>. The project PID2020-115310RB-I00 is funded by Spanish Ministry of Science and Innovation <https://www.aei.gob.es/convocatorias/buscador-convocatorias/proyectos-idi-2020-modalidades-retos-investigacion-generacion>. Funding for open access charge: Universidad de Málaga/CBUA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Funding Funding for open access publishing: Universidad de Málaga/CBUA.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Araci, D. (2019). FinBERT: Financial sentiment analysis with pre-trained language models. <https://arxiv.org/abs/1908.10063>
- Bauman, R. (2001). The ethnography of genre in a Mexican market: Form, function, variation. In P. Eckert & J. Rickford (Eds.), *Style and sociolinguistic variation* (pp. 55–77). Cambridge University Press.
- Bednarek, M., & Caple, H. (2017). *The discourse of news values: How news organizations create newsworthiness*. Oxford University Press.
- Bhatia, V. K. (2004). *Worlds of written discourse*. Continuum.
- Bhatia, V. K. (2023). Critical genre analysis of professional discourse. En *The Routledge Handbook of Discourse Analysis* (pp. 358–372). Routledge.
- Biber, D., Eniko Csomay, Jones, J. K., & Keck, C. M. (2004). *A corpus linguistic investigation of vocabulary-based discourse units in university registers*. 53–72. https://doi.org/10.1163/9789004333772_005
- Braud, C., & Denis, P. (2016). Learning connective-based word representations for implicit discourse relation identification. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 203–213.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & others. (2020). Language models are few-shot learners. *arXiv Preprint arXiv:2005.14165*.
- Cenni, I. (2024). Sharing travel experiences on TripAdvisor: A genre analysis of negative hotel reviews written in French, Spanish and Italian. *Journal of Pragmatics*, 221, 76–88. <https://doi.org/10.1016/j.pragma.2023.12.015>
- Chalkidis, I., Fergadiotis, M., Kotitsas, S., Malakasiotis, P., Aletras, N., & Androutsopoulos, I. (2020). An empirical study on large-scale multi-label text classification including few and zero-shot labels. In *Proc. 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) 7503–7515* (Association for Computational Linguistics, 2020).
- Coppock, A., Ekins, E., & Kirby, D. (2018). The long-lasting effects of newspaper op-eds on public opinion. *Quarterly Journal of Political Science*, 13(1), 59–87. <https://doi.org/10.1561/100.00016112>
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95–113. <https://doi.org/10.1016/j.pragma.2007.10.003>
- Downing, A. (2001). Thematic progression as a functional resource in analysing texts. *CLAC (Circulo De Lingüística Aplicada a La Comunicación)*, 5(2), 23–42.
- Egbert, J., Wizner, S., Keller, D., Biber, D., McEnery, T., & Baker, P. (2021). Identifying and describing functional discourse units in the BNC Spoken 2014. *Text & Talk*, 41(5–6), 715–737. <https://doi.org/10.1515/text-2020-0053>
- Fairclough, N. (2015). *Language and Power (Third edition)*. Routledge.
- Fernández-Cruz, J., & Moreno-Ortiz, A. (2023). Tracking diachronic sentiment change of economic terms in times of crisis: Connotative fluctuations of ‘inflation’ in the news discourse. *Plos One*, 18(11), e0287688.
- Fernández-Cruz, J., & Moreno-Ortiz, A. (2024). *Economic terms in the news during the Great Recession: A diachronic sentiment and collocational analysis (1st, New ed ed.)*. Peter Lang. <https://doi.org/10.3726/b21287>

- Finlayson, M. A., & Erjavec, T. (2017). Overview of Annotation Creation: Processes and Tools. In N. Ide & J. Pustejovsky (Eds.), *Handbook of linguistic annotation* (pp. 167–191). Springer Netherlands. https://doi.org/10.1007/978-94-024-0881-2_5
- Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., & Smith, N. A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. *arXiv Preprint arXiv:2004.10964*.
- Halliday, M. A. K. (2014). *Introduction to Functional Linguistics (4th Edition)*. Routledge.
- Hyland, K. (2018). Genre and discourse analysis in language for specific purposes. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (1st ed., pp. 1–8). Wiley. <https://doi.org/10.1002/9781405198431.wbeal0452.pub2>
- Kehoe, A., & Gee, M. (2007). New corpora from the web: Making web text more 'text-like'. *Studies in variation, contacts and change in English*, 2. https://varieng.helsinki.fi/series/volumes/02/kehoe_gee/
- Kolhatkar, V., Wu, H., Cavasso, L., Francis, E., Shukla, K., & Taboada, M. (2020). The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments. *Corpus Pragmatics*, 4(2), 155–190. <https://doi.org/10.1007/s41701-019-00065-w>
- Kristeva, J. (1986). Word, dialogue and novel. In T. Moi (Ed.), *The Kristeva reader*. Basil Blackwell.
- Jones, R. H. (2024). *Discourse analysis* (Third edition). Routledge, Taylor & Francis Group. <https://doi.org/10.4324/9781003377405>
- Labov, W., & Waletzky, J. (1967). Narrative analysis. In J. Helm & J. H. MacNeish (Eds.), *Essays on the verbal and visual arts* (pp. 12–44). University of Washington Press Seattle.
- Liu, B. (2015). *Sentiment Analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Mann, W. C., Matthiessen, C. M. I. M., & Thompson, S. A. (1992). Rhetorical Structure Theory and text analysis. In W. C. Mann & S. A. Thompson (Eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text* (pp. 39–78). John Benjamins.
- Melo, J. M. D., & Assis, F. D. (2016). Gêneros e formatos jornalísticos: Um modelo classificatório. *Intercom: Revista Brasileira de Ciências Da Comunicação*, 39(1), 39–56. <https://doi.org/10.1590/1809-5844201613>
- Milutinović, I. (2021). Media framing of COVID-19 pandemic in the transitional regime of Serbia: Exploring discourses and strategies. *Media, Culture & Society*, 43(7), 1311–1327. <https://doi.org/10.1177/0163443720986003>
- Montani, I., & Honnibal, M. (2018). Prodigy: A new annotation tool for radically efficient machine teaching. *Artificial Intelligence, to appear*.
- Montani, I., Honnibal, M., Van Landeghem, S., Boyd, A., Peters, H., Samsonov, M., Geovedi, J., McCann, P. O., Regan, J., Orosz, G., Altinok, D., Kristiansen, S. L., Roman, Fiedler, L., Howard, G., Wannaphong Phatthiyaphaibun, Tamura, Y., Explosion Bot, Bozek, S., ... Henry, W. (2021). *Prodigy v1.11.4* (Version v3.1.0) [Computer software]. Explosion. <https://doi.org/10.5281/ZENODO.1212303>
- Nie, A., Bennett, E., & Goodman, N. (2019). DisSent: Learning sentence representations from explicit discourse relations. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4497–4510.
- Oberländer, L. A. M., Kim, E., & Klinger, R. (2020). GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, 1554–1566.
- Orts Llopis, M. A. (2015). Traduciendo al oráculo: Desafíos traductológicos en la obra de Paul Krugman. Current Approaches to Business and Institutional Translation: Proceedings Of the International Conference Economic, Business, Financial and Institutional Translation, 147–158.
- Orts Llopis, M. A. (2017). Krugman vs Garicano: Individual and cultural differences in the rhetoric of two economic op-ed writers. *HERMES - Journal of Language and Communication in Business*, 55, 121–141. <https://doi.org/10.7146/hjlc.v0i55.24293>
- Pascual, D. (2023). Scientific dissemination and professional practices through digital media: The study of pragmatic strategies in the communication of international research projects [PhD Dissertation, Universidad de Zaragoza]. <https://zaguan.unizar.es/record/126892>
- Pak, I., & Teh, P. L. (2018). Text segmentation techniques: a critical review. *Innovative Computing, Optimization and Its Applications: Modelling and Simulations*, 167–181.

- Prasad, R., Webber, B., & Joshi, A. (2017). The Penn Discourse Treebank: an annotated corpus of discourse relations. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation* (pp. 1197–1217). Springer Netherlands. https://doi.org/10.1007/978-94-024-0881-2_45
- Quasthoff, U., Heller, V., & Morek, M. (2017). On the sequential organization and genre-orientation of discourse units in interaction: An analytic framework. *Discourse Studies*, 19(1), 84–110. <https://doi.org/10.1177/14614456166835>
- Revilla Vicente, R. (2008). EAP, Business English and Swales' approach to genre analysis. *ES: Revista de Filología Inglesa*, 29, 153–165.
- Rocci, A. (2014). The discourse system of financial communication. *Cahiers Du Centre De Linguistique Et Des Sciences Du Langage*, 34, 201–221.
- Roslyng, M. M., & Dindler, C. (2023). Media power and politics in framing and discourse theory. *Communication Theory*, 33(1), 11–20. <https://doi.org/10.1093/ct/qtac012>
- Sacks, H. (1995). *Lectures on conversation I & II*. Blackwell.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.
- Sileo, D., Van de Cruys, T., Pradel, C., & Muller, P. (2020). DiscSense: Automated semantic analysis of discourse markers. In N. Calzolari et al. (Eds.), *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 991–999). European Language Resources Association. <https://aclanthology.org/2020.lrec-1.125>
- Sinclair, J. (2004). *Trust the text*. Routledge.
- Spencer, L. (2005). *A step-by-step guide to persuasive writing*. The Rosen Publishing Group.
- Swales, J. (1981). *Aspects of article introductions*. University of Aston.
- Swales, J. (1990). *Genre analysis: English for academic and research settings*. Cambridge University Press.
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.
- Upton, T. A., & Cohen, M. A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, 11(5), 585–605.
- Vilar, H., & Tolchinsky, L. (2022). The rhetorical structure of analytical writing: A developmental approach. *Text & Talk*, 42(1), 131–152. <https://doi.org/10.1515/text-2019-0228>
- Waring, H. Z. (2018). *Discourse analysis*. Routledge.
- Weisser, M. (2019). The DART annotation scheme: Form, applicability & application. *Studia Neophilologica*, 91(2), 131–153. <https://doi.org/10.1080/00393274.2019.1616218>
- Zheng, S. (2014). A critical discourse analysis of financial remarks: A case study. *International Journal of English Linguistics*, 4(5), Article p105. <https://doi.org/10.5539/ijel.v4n5p105>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.