Information Exchange vs. Opinion Discussion as Distinct Communicative Contexts in Online Conversations

A Corpus-Based Study of Speech Acts and Rhetorical Relations Dynamics in Dialogue

Gonzalo Freijedo Aduna
Institut Jean Nicod
CNRS/ENS/EHESS/PSL University
gfreijedoaduna@ens.psl.eu

Alda Mari
Institut Jean Nicod
CNRS/ENS/EHESS/PSL University
alda.mari@ens.com

Abstract

We present a framework for analyzing online conversations that distinguishes communicative contexts through fine-grained annotation of speech acts and rhetorical relations. Using a segmented Reddit corpus, we compare information-seeking and opinion-driven exchanges, showing that these contexts exhibit distinct discourse structures. Opinion discussions display more frequent markers of agreement, disagreement, and biased questions, reflecting overt negotiation of viewpoints. In contrast, information-oriented dialogues emphasize clarification and elaboration. Our findings highlight how communicative context shapes the rhetorical dynamics of dialogue and offer a resource for modeling common ground in asynchronous conversation.

1 Introduction

Distinguishing between communicative contexts in conversation provides a productive lens for analyzing how information—and misinformation—is exchanged, evaluated, and negotiated in online discourse. In this paper, we focus on two broad types of communicative contexts: information-seeking or task-oriented exchanges, and opinion discussion-oriented interactions. Understanding how these distinct contexts shape discourse structure and common ground management allows us to better characterize the dynamics of epistemic alignment, disagreement, and clarification—key processes in the spread or containment of (mis)information in digital environments.

To analyze how these contexts shape conversational dynamics, we introduce an annotation scheme that captures both the speech acts (Searle, 1969) performed in individual segments and the rhetorical relations (Asher and Lascarides, 2003) that connect them. Both layers are essential for a fine-grained discourse analysis. However, the rhetorical structure—how utterances relate to one another—offers distinctive insight into context-sensitive patterns of information flow and engagement, particularly in relation to the construction and negotiation of common ground.

Central to our contextual distinction is a more precise treatment of *opinion*, which diverges from its conventional usage in NLP—especially in sentiment analysis and stance detection—where it is typically linked to affective polarity (positive, negative, or neutral). Instead, we adopt an epistemic perspective that distinguishes between utterances asserting *facts*, or verifiable claims about the external world, and those expressing *opinions*, understood as subjective beliefs, preferences, or evaluative judgments. This contrast, grounded in the semantic literature (Saurí and Pustejovsky, 2009; Portner, 2009; Giannakidou and Mari, 2021b), informs the design of our annotation schema, which aims to capture conversations dynamics rather than emotional tone or stance of individual utterances.

Accordingly, the Speech Act taxonomy introduced in Section 4 encodes distinctions relevant to the type of information conveyed, while the set of rhetorical relations models how utterances contribute to the evolving discourse structure.

Our work also intersects with research in discourse parsing and argument mining (Joty and Mohiuddin, 2018; Habernal and Gurevych, 2017; Xing and Carenini, 2021), which has aimed to map the structure of online conversations through relations. However, while many such approaches rely on supervised classification of argument structure or coherence relations, we focus on fine-grained, linguistically motivated units—speech acts—and their rhetorical relations to one another. This allows us to capture the dynamic, interactional dimension of discourse that reflects participants' coordination and negotiation over shared content.

To do so, we segment posts and comments into elementary utterances using *Stanza* (Qi et al., 2020), combined with rule-based lexical cues, and

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

identify the specific utterance each one responds to—what we refer to as the reply-to chain. This reconstruction of adjacency pairs (Sacks et al., 1978) allows us to analyze how dialogue structure develops at the level of individual speech acts.

HYP. We hypothesize that the type of conversational context—whether the dialogue addresses opinions or factual matters—influences the distribution of speech acts and rhetorical relations. We expect this to result in distinct patterns characteristic of each context. Our analysis supports this hypothesis, revealing the following trends:

- Explicit agreement and disagreement are markedly more frequent in opinion-oriented discussions, suggesting that such contexts involve heightened negotiation of subjective viewpoints and more overt contributions to the common ground.
- In contrast, clarification and elaboration requests are more frequent in fact-oriented contexts, indicating a greater emphasis on reducing ambiguity and refining shared knowledge.
- We also observe a marked asymmetry in the distribution of rhetorical questions—understood here as questions that implicitly convey information—which are far more prevalent in opinion-driven exchanges.

Finally, we examine the distribution of rhetorical relations within individual comments, shedding light on the internal argumentative structure of contributions.

The paper is structured as follow. Section 2 introduces the pragmatic concepts underlying our annotation scheme. Related work is reviewed in Section 3. In Section 4, we describe the annotation process in detail, and Section 5 presents the main findings. We conclude in Section 6.

2 Theoretical Foundations

As discussed in the Introduction, we examine online conversations through three interconnected concepts: **speech acts**, **rhetorical relations**, and **common ground**. Specifically, we investigate how interactions between speech acts—interpreted through rhetorical relations—contribute to the construction of common ground across different communicative contexts.

Speech Acts By speech acts, we follow Searle (1969) in referring to the type of action a speaker performs in speaking: asserting, questioning, advising, commanding, expressing emotions, etc. (Austin, 1962; Levinson, 1983; Clark, 1996)[1]. These actions typically elicit specific kinds of responses—agreement, elaboration, rejection, or silence. Action and reaction form an *adjacency pair* (Sacks et al., 1978), revealing an underlying expectation structure between conversational contributions. In this view, speech acts are not isolated; they function in sequence, and their meaning unfolds in the context of their relations to prior and subsequent acts.

Rhetorical Relations These inter-utterance links are captured by rhetorical relations. Drawing from Segmented Discourse Representation Theory (SDRT; (Asher and Lascarides, 2003, 2008), we understand rhetorical relations as the organizing principles that ensure coherence across a discourse. Now, rhetorical relations can be either **monological** or **dialogical**. In the monological sense, they structure the internal logic of a single speaker's discourse, connecting propositions through relations such as EXPLANATION, ELABORATION, or CONTRAST. In the dialogical sense, they function similarly to adjacency pairs, linking the speech acts of one speaker to those of another—e.g., a question followed by an answer, or a claim followed by a challenge. In this way, rhetorical relations mediate the interaction between speakers, revealing how conversational moves either advance, negotiate, or contest shared content.

We thus view the rhetorical organization of dialogue as the mechanism through which the common ground is dynamically built. Our annotation framework captures this by labeling both the illocutionary force of each utterance (its speech act type) and its rhetorical relation to another utterance—whether it agrees, challenges, clarifies, elaborates, or performs another dialogic function.

Common Ground The notion of common ground was first defined by Stalnaker (1978) as the set of propositions believed to be shared by the participants in a conversation[2]. In Stalnaker's view, the content and form of an utterance depend in part on what is assumed to be in the common

[1] See Portner (2018) for an overview.

[2] Yalcin (2024) has argued that the epistemic attitude at the basis of common ground is knowledge about what is commonly accepted.

ground. However, in online asynchronous conversations—such as those on Reddit—participants are typically unfamiliar with each other. This raises the question: what content can be presumed to be shared?

The issue of unfamiliarity is addressed by Semeijn (2024), who argues that, in the context of non-face-to-face communication such as the one under discussion, we should conceive of the speakers' beliefs as beliefs de dicto rather than de re. That is, beliefs not about the beliefs of particular individuals, but of the members of a group or community as such—in other words, beliefs that are presupposed to be held simply by virtue of belonging to that community. A closely related perspective is found in Breitholtz (2020), who uses the notion of *topoi* to model how common-sense reasoning and shared inferential patterns structure conversational contributions in dialogue, extending the KoS framework (Ginzburg, 2012) to account for such mechanisms.

Our focus, however, is not on what is presupposed to be shared, but rather on how content is added to and negotiated within the common ground—a process known as *grounding*, which, at Level 4 in Clark and Schaefer (1989), corresponds to the mutual belief of a proposition by interlocutors. This is the central insight of a series of refinements to Stalnaker's model. Scholars such as Lewis (1979), Ginzburg (1996, 2012), Farkas and Bruce (2010), Krifka (2015) and Geurts (2024) propose dynamic models—e.g., the *conversational scoreboard*, *dialogue gameboard*, and *Table model*—which explicitly represent how conversational moves introduce propositions into a shared space of commitments and expectations. These frameworks conceptualize interaction as a process of coordinated negotiation, in which each speech act proposes a potential update to the common ground.

We adopt this perspective in hypothesizing that the structure of negotiation spaces varies across communicative contexts. In opinion-oriented discussions, we expect more frequent and explicit moves of alignment, disagreement, or challenge, leading to a denser web of rhetorical interactions. In contrast, in task- or information-oriented contexts, we expect speech acts to function more linearly—focused on information verification, clarification, and implicit agreement. In both cases, it is through the sequence and structure of speech acts and rhetorical relations that the common ground is constructed.

3 Related Work

While, to the best of our knowledge, no existing work interrelates the three pragmatic notions we explore in this paper in the way we do, several proposals are closely aligned with our objectives.

A first line of relevant research addresses how different types of assertions interact with the common ground. In this regard, Beltrama (2018) presents two experimental studies investigating how subjective and objective assertions—corresponding to our categories of ASSESSMENT and ASSERTION, respectively, as outlined in Section 4.2—are integrated into the common ground. In one of the experiments, native English speakers were asked whether a proposition p, conveyed through either an ASSERTION or an ASSESSMENT, would be incorporated into the common ground in the absence of an explicit confirmation by the addressee[3]. The findings show that failing to explicitly accept p is more often interpreted as implicit acceptance following ASSERTIONS than ASSESSMENTS, indicating that propositions involving subjective predicates are less likely to be accommodated by default. Our results, presented in Section 5, complement these findings.

In terms of annotated discourse data, the STAC corpus (Asher et al., 2016; Afantenos et al., 2015) represents a key effort in modeling interactional structure. It consists of multi-party chat dialogues annotated for discourse structure within the framework of Segmented Discourse Representation Theory (SDRT). STAC aims to capture how participants navigate conflicting goals in competitive conversational settings, such as online games.

Another relevant contribution is Amidei et al. (2021), who present QTMM2012c+, a corpus of *Moral Maze* debates annotated following Inference Anchoring Theory (IAT), which captures the structure of argumentative discourse through transitions between locutions, propositional relations, and illocutionary connections linking speech acts to their

[3]This corresponds to the 'Continued Attention' type of evidence of undersanding in Clark and Schaefer (1989), which, in face-to-face communication is one of "the three most common forms of positive evidence" of understating (*loc. cit.*, p. 131). See also Roque and Traum (2008) for more details on degrees of grounding in task-oriented dialogue settings. However, in online communication—our study's focus—the absence of an explicit reply is typically taken as a lack of engagement in the conversation, since many nonverbal grounding cues are missing.

propositional content. These propositional relations correspond, roughly, to what we describe as rhetorical relations. While their focus is on modeling argument structure and the flow of debate rather than common ground dynamics *per se*, their work highlights the value of discourse-level representations for analyzing dialogic interaction—an approach we adopt in our study.

Other approaches have applied discourse-based analysis to monologic texts. Asher et al. (2009), for instance, propose an annotation schema for opinion expressions that combines a detailed lexical semantic typology with discourse-level analysis. Their framework integrates rhetorical relations—such as SUPPORT, CONTRAST, and RESULT—defined also within SDRT, to model how discourse structure modulates the strength and polarity of opinions. Unlike our approach, which targets the interactive dynamics of multi-party conversations, their analysis is centered on isolated commentaries or reviews.

Similarly, Trnavac and Taboada (Trnavac and Taboada, 2012) examine how nonveridical rhetorical relations—such as concession and condition—contribute to the evaluative layer of discourse, allowing speakers to position themselves subtly with respect to propositions. These insights support our emphasis on rhetorical relations as carriers of both informational and attitudinal content, particularly in opinion-driven exchanges.

Finally, several recent projects have used manual annotation to investigate speech act variation in isolated messages, rather than in conversational sequences. For example, Jegede (2025) analyze speech act distributions across platforms, distinguishing between public and private contexts. Laurenti et al. (2022a,b) classify speech acts in crisis communication to support emergency response, treating each message as an independent unit.

3.1 NLP approaches to automatic detection of conversation structure

As indicated in the Introduction, our goal is to extract conversational structure at the level of elementary speech acts in order to analyze how common ground is constructed across distinct communicative contexts. This task intersects with ongoing research in NLP on the automatic identification of conversational dependencies, typically addressed through unsupervised or semi-supervised methods. Joty and Mohiuddin (2018) and Joty et al. (2013) propose hybrid models for recognizing speech acts and their dependencies in asynchronous conversations, such as forums and emails, combining sentence-level features with conversational context. Similarly, Xing and Carenini (2021) introduces a method for topic segmentation based on coherence scoring between utterance pairs. However, these approaches generally target coarse-grained conversational turns and do not account for the rhetorical dependencies between fine-grained speech acts.

Thread disentanglement has been widely studied in the context of online forums and multi-party dialogues, where conversational structure is often obscured by interface constraints or asynchronous posting. Prior work has proposed both heuristic and machine learning approaches to infer reply-to links, using features such as lexical overlap, temporal proximity, speaker turns, or sequential modeling (Mehri and Carenini, 2017; Kummerfeld et al., 2019; Qiu et al., 2020). While these methods provide valuable insights, they are typically applied to unsegmented messages or full turns. In contrast, our task focuses on elementary discourse units—individual speech acts within segmented comments—and the rhetorical dependencies between them.

Given this complexity, we manually corrected the reply-to chains during annotation to ensure that each utterance is linked to the specific speech act it responds to, rather than simply inheriting the parent comment's position. This enables a more accurate representation of interactional structure than is possible using platform metadata alone. In line with recent calls for thoughtful data design in multi-party dialogue research (Mahajan and Shaikh, 2021), the resulting dataset will support future development and evaluation of disentanglement models adapted to segmented, annotated discourse.

4 Corpus and Annotation process

To study discourse patterns in the management of common ground across different conversational settings, we systematically compiled a dataset by scraping six Reddit communities over a period spanning from October 26, 2024, to February 28, 2025. We grouped these into two communicative contexts: the **Info Group**, oriented toward information-seeking and task-focused exchanges, includes *askscience*, *DoItYourself*, and *todayilearned*; the **Opinion Group**, centered on opinion-sharing and evaluative dialogue, includes *AskPolitics*, *changemyview*, and *prochoice*. After segmentation (see Section 4.1), the Info Group ac-

counts for 48.66% of the annotated segments, while the Opinion Group comprises 51.35%, ensuring a relatively balanced distribution across communicative contexts.

4.1 Segmentation

As outlined in the Introduction, we decomposed each comment into smaller discourse units corresponding to individual speech acts. We refer to these units as segments, or utterances in the sense of Popescu-Belis (2007): "parts of a turn that accomplish an elementary dialogue function." We then reorganized the reply-to chain so that each segment is explicitly linked to the specific other segment it responds to. This restructuring transforms the original tree-like conversation structure into a graph (see Figure 1b, where square colors represent speech act types and arrow colors indicate rhetorical relations).

For preprocessing, we used *Stanza* (Qi et al., 2020) to perform tokenization, POS tagging, dependency parsing and sentence segmentation. We then applied additional rule-based segmentation to extract finer-grained utterances. This involved identifying discourse markers (e.g., *but*, *because*, *although*), compound markers (e.g., *even though*, *so that*), and leveraging syntactic cues such as the presence of a subject and predicate to avoid splitting coordinated verb phrases. Each resulting segment was assigned a unique identifier derived from the original comment ID. The occasional errors in the segmentation were corrected during the annotation process.

To ensure the reliability of the annotation scheme, the first 500 segments were independently annotated by two PhD students. This process included two preliminary rounds—one with 50 annotations and another with 150—which served to refine the guidelines and resolve initial discrepancies. Following this calibration phase, the annotators independently annotated an additional 300 segments using the finalized scheme, achieving an inter-annotator agreement of **Cohen's Kappa = 0.75**.

The remaining annotation was then completed on 1,483 new instances. In total, the corpus comprises 49 posts, including 607 commentaries (original posts and replies), segmented into 1,983 speech acts (see Table 1 for details).

We set two levels of annotations for each segment: its speech act category, and its rhetorical relation, that is, the way it relates to the speech act that is replying to.

Subreddit	Posts	Comments	Segments
Askpolitics	3	40	143
DIY	8	79	244
Askscience	11	112	427
Changemyview	9	115	575
Prochoice	5	101	300
Todayilearned	13	160	294
Total	**49**	**607**	**1983**

Table 1: Dataset statistics per subreddit.

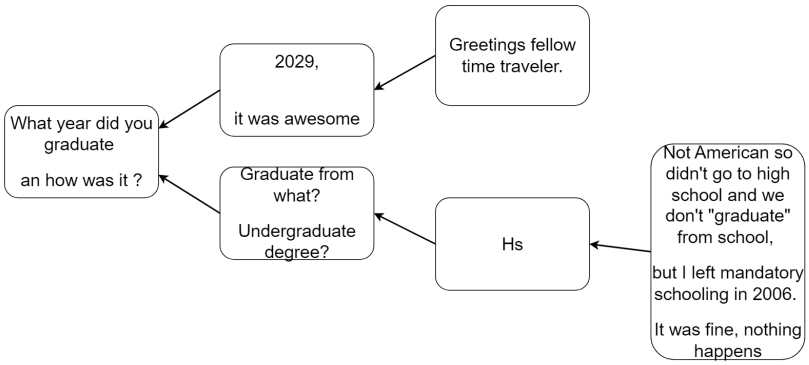
4.2 Speech Acts

Our categorization of speech acts is closely tied to the goal of the annotation. Given the need to distinguish between conversational contexts, we subdivided **assertions** and **questions** into more specific categories[4]:

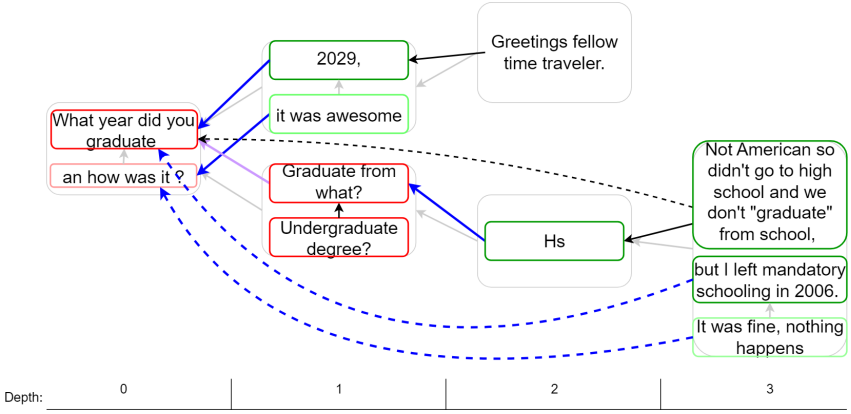
- ASSESSMENT: statements involving subjectivity, evaluation, or non-factive modality (deontic or epistemic).
- ASSERTION: statements conveying information about the world, even if false.
- QUESTION (OPINION): questions about subjective values or belief attitudes. These may be either *straight* (genuinely seeking an answer) or *biased* (framed to express a stance or lead the interlocutor to a preferred conclusion).
- QUESTION (FACTS): questions that seek information about objective states of the world. Like their opinion-based counterparts, they may appear in either a *straight* or a *biased* form.
- ADVISE (RECOMMENDATION/SUGGESTION): typically imperatives proposing actions or cognitive strategies.
- EXPRESSIVE: expressions of emotion, humor, or interpersonal stance.
- OTHER: Expressions that do not fall into any of the categories defined above.

Biased questions merit special attention, as they often convey information rather than seek it. Given their assertive force, several scholars have proposed analyzing them as a type of declarative—either as assessments or assertions—rather than as genuine interrogatives (see for the foundational discussions However, to the extent that they retain some interrogative force, and some degree of uncertainty, we chose to annotate them as this special kind of

[4]Here we follow Beltrama (2018), who argues that "subjective predicates display different discourse behavior from objective assertions", and, as a consequence, subjective assertions "should be assigned a special illocutionary profile".



(a) Tree structure of unsegmented commentaries



(b) Graph structure of segmented commentaries

Figure 1: Comparison of structures before and after segmentation.

question. We will report the results of this election in 5.1.1.

We refer the reader to Appendix A for illustrative examples of each speech act type in our corpus.

4.3 Rhetorical Relations: Inter- and Intra-comment Structure

Rhetorical relations in our framework correspond to reactions to speech acts, and serve to show how the common ground is constructed and negotiated. We distinguish between two types of rhetorical structure: *inter-comment* (dialogical) relations, which link segments of distinct comments in a thread, and *intra-comment* (monological) relations, which organize discourse segments within a single comment.

Inter-comment (dialogical) relations

- AGREEMENT: Indicates that the speaker aligns with the previous speech act, reinforcing shared beliefs or viewpoints.
- DISAGREEMENT: Signals a rejection or challenge to the previous speaker's proposition, often leading to negotiation or revision of the common ground.
- ANSWER: A direct response to a previous question or query, providing information or clarification.
- REQUEST CLARIFICATION/ELABORATION: Seeks additional details or clarification about a prior statement or question.
- CONTINUATION -EX-: Indicates that the current speech act is a continuation or elaboration of a segment from a different comment. In some cases, it may implicitly express agreement by expanding on a shared point.
- OFF-TOPIC: Introduces content that deviates from the current discussion.

Intra-comment (monological) relations

- CONTINUATION -IN-: Indicates that the current speech act continues or develops a segment within the same comment. Following Asher et al. (2009), we subdivide this relation into:
  - EXPLANATION/ELABORATION (SUPPORT): Clarifies, expands, or justifies the preceding segment.
  - CONTRASTIVE: Introduces an alternative or opposition to the content of the first segment.
  - RESULT: Presents an outcome or consequence that follows from the previous segment.
  - CONTINUATION: Extends or completes a previous segment without introducing a new argumentative relation.

We refer the reader to Appendix B for illustrative examples of each rhetorical relation in our corpus.

5 Results

The presentation of our results is divided into two subsections. We first examine inter-comment rhetorical relations (5.1), followed by an analysis of intra-comment rhetorical relations (5.2). In both cases, we investigate whether distinct patterns of rhetorical relations emerge depending on the communicative context.

5.1 Inter-comment rhetorical relations

Given our goal of examining how common ground is managed across communicative contexts, inter-comment rhetorical relations are arguably the most crucial. To ensure the robustness of our findings, we define communicative context in two complementary ways: globally (5.1.1), by treating entire communities as context, and locally (5.1.2), by considering the immediately preceding segment as context.

5.1.1 Communities as Communicative Contexts

We begin by approaching the notion of context through the subreddit—or community—in which a conversation takes place. Our corpus design reflects this assumption: we selected three subreddits characterized by information-seeking or task-oriented goals (*askscience*, *todayilearned*, and *DIY*) and three centered on opinion exchange and evaluative discussion (*changemyview*, *Askpolitics*, and *prochoice*).

To empirically validate this, we first analyzed the distribution of speech act types in each subreddit, focusing on the contrast between ASSERTION and ASSESSMENT. Table 2 reports the raw frequencies for each category. To assess whether the prevalence of factual versus opinion-based assertions differed significantly within each subreddit, we then conducted two-sided binomial tests.

The results are consistent with our classification of communicative contexts into two groups: an **Info Group** (*askscience*, *todayilearned*, *DIY*) and an **Opinion Group** (*changemyview*, *Askpolitics*, *prochoice*).

In the **Info Group**, factual assertions clearly predominate[5]. In *askscience*, ASSERTION significantly outweighs ASSESSMENT (287 vs. 53, $\chi^2 = 67.05$, $p < 0.001$); *todayilearned* shows a similar trend (151 vs. 73, $\chi^2 = 10.23$, $p < 0.01$); and *DIY* likewise favors factual over opinion-based assertions (89 vs. 46, $\chi^2 = 7.95$, $p < 0.01$).

Conversely, in the **Opinion Group**, the distribution is reversed. In *changemyview*, opinion-based assertions vastly outnumber factual ones (427 vs. 82, $p < 0.001$), in line with its focus on belief revision and evaluative discourse. *Askpolitics* shows a marked preference for opinions (85 vs. 35, $p < 0.001$), and *prochoice* similarly favors assessments over assertions (153 vs. 57, $p < 0.001$). These findings confirm that the distribution of epistemic speech acts aligns with the intended communicative function of each community[6].

Given this typological distinction, we now turn to the distribution of inter-comment rhetorical relations across groups. As shown in Figure 2 and Table 3, the **Opinion Group** accounts for 66.4% of all instances of AGREEMENT/DISAGREEMENT, compared to 33.6% in the **Info Group**.

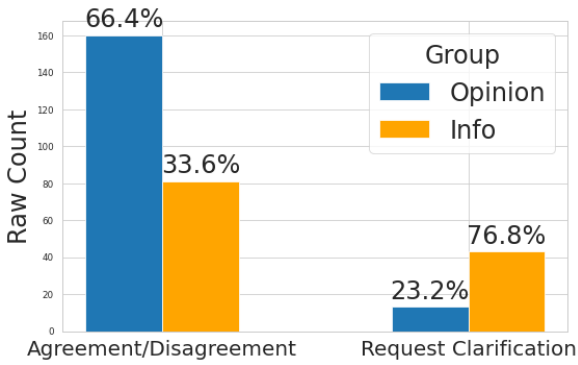


Figure 2: Distribution of Agreements/Disagreements and Clarification Requests

This suggests that in the **Opinion Group**, discussions involve more explicit negotiation of common ground, with participants frequently expressing agreement or disagreement. On the other hand, the **Info Group** appears to have a more informational and cooperative approach to common ground management. The higher proportion of requests for clarification suggests that users engage in epistemic alignment—seeking to refine, verify, or expand on information rather than debating positions.

While the categories of ASSERTION and ASSESSMENT offer insight into dominant speech act patterns across communities, they do not exhaust the

[5]In what follows, significance levels were calculated using the chi-square test of independence, based on 2×2 contingency tables comparing the frequency of ASSERTION and ASSESSMENT speech acts versus all other speech acts within each subreddit.

[6]A similar epistemic pattern holds for interrogative acts.

Subreddit	Assertion	Assessment	Q (fact)	Q (op.)	Expr.	Adv/Rec	Other	Total
DIY	89	46	31	2	26	47	3	244
askscience	287	53	61	4	11	11	0	427
todayilearned	151	73	23	6	32	8	1	294
Askpolitics	35	85	0	14	4	4	1	143
changemyview	82	427	7	28	17	8	6	575
prochoice	57	153	3	26	48	12	1	300
Total	701	837	125	80	138	90	12	**1983**

Table 2: Speech Act Distribution Across Subreddits

Group	Agr./Disagr.	Ans.	Cont.-EX	Cont.-IN	OffT.	Clarif.
Info	81	76	130	579	24	43
Opinion	160	19	86	600	5	13

Table 3: Distribution of rhetorical relations by group. The difference in AGREEMENT/DISAGREEMENT is significant ($\chi^2(1, N = 1816) = 34.30, p < 0.001$).

pragmatic functions found in these conversations. One particularly revealing case involves *biased questions*—interrogatives that, despite their syntactic form, function primarily to convey an evaluative stance or presuppose a particular answer. As discussed in Section 4.2, these acts blur the line between questioning and asserting.

To examine how their use varies across communicative contexts, Figure 3 shows the distribution of biased questions in both the **Info Group** and the **Opinion Group**.

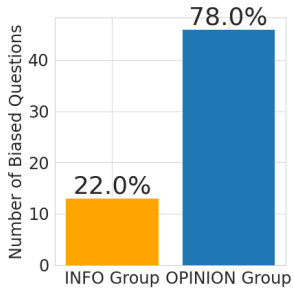


Figure 3: Distribution of biased questions across global communicative contexts (Info vs. Opinion).

The distribution confirms that biased questions are far more prevalent in opinion-oriented contexts, which seems to be in line with the findings reported in Amidei et al. (2021), where assertive (biased) and rhetorical questions outnumber pure (straight) questions in an opinion-discussion setting[7]. This also aligns with the common wisdom since Ladd (1981)[8] that the implied proposition in a biased question retains a degree of epistemic uncertainty. Taken together, these findings suggest that—despite their assertive flavor—biased questions function pragmatically more like ASSESSMENTS than ASSERTIONS, as they convey evaluative stance rather than factual commitment.

It is important to note that the proportions reported here reflect rhetorical relations across groups, *independently* of the speech acts they respond to. To examine how these relations function within their immediate context, we now turn to their interaction with local speech act sequences.

5.1.2 Local communicative context

A second way to define communicative context is by looking at the immediate conversational environment—specifically, the speech act to which a given segment responds. Analyzing these local reply structures reveals the interactional tendencies associated with each act type. This structural perspective abstracts away from topical content, focusing instead on how different contributions (e.g., assertions, assessments) organize and shape the flow of dialogue across contexts.

This local perspective reveals patterns similar to those observed when context is defined at the community level. Among all AGREEMENT/DISAGREEMENT relations, 73.3% are directed at ASSESSMENTS, while only 26.7% target

[7]Note, however, that their analysis is based on an oral corpus collected in a highly structured, interview-style setting, whereas our data comes from online forum interactions which, while subject to community guidelines, allow for more varied and less formally constrained contributions.

[8]See Larrivée and Mari (2022); Giannakidou and Mari (2021a) for recent discussions and references.

Subreddit	Continuation	Expl./Elab.	Contrastive	Result
Askpolitics	61.46%	20.83%	11.46%	6.25%
DIY	75.82%	12.42%	5.88%	5.88%
Askscience	58.76%	25.43%	11.00%	4.81%
Changemyview	67.98%	19.21%	9.61%	3.20%
Prochoice	86.24%	10.58%	1.59%	1.59%
Todayilearned	75.56%	14.07%	8.15%	2.22%

Table 4: Distribution (in %) of intra-rhetorical relations within each subreddit. Percentages are relative to the total number of intra-rhetorical relations in the subreddit.

ASSERTIONS, reinforcing the association between evaluative discourse and expressions of stance. This distribution is statistically significant ($\chi^2(1, N = 1816) = 34.76$, $p < 0.001$), suggesting that opinion-based speech acts are considerably more likely to elicit reactions of agreement or disagreement. In contrast, 67.4% of REQUEST CLARIFICATION relations respond to ASSERTIONS, and only 32.6% to ASSESSMENTS, indicating that clarification is more often sought in response to informational content. This difference is also statistically significant ($\chi^2(1, N = 1816) = 7.38$, $p < 0.01$).

5.2 Intra-comment Rhetorical Relations

We conclude by examining intra-comment rhetorical relations to understand how users structure their contributions within individual comments. While this level of analysis does not draw a sharp line between information-oriented and opinion-oriented contexts, it still reveals notable stylistic differences across communities (Table 4).

Among the four relations considered, CONTINUATION is the least informative, often reflecting linear progression without rhetorical segmentation. In contrast, EXPLANATION/ELABORATION, CONTRASTIVE, and RESULT signal greater internal organization, suggesting justifications, distinctions, or inferences—markers of structured discourse.

Subreddits like *askscience* and *changemyview* show higher proportions of these structuring relations, consistent with more elaborated and argumentative contributions. By contrast, in *prochoice*, where monologic discourse is more frequent, CONTINUATION overwhelmingly dominates, indicating a simpler, more linear style.

Overall, while intra-comment structure does not fully align with our broader contextual typology, the relative presence of rhetorically rich relations serves as a proxy for discursive complexity within communities.

6 Conclusion

This work explores how rhetorical relations shape the construction and negotiation of common ground in both opinion-driven and fact-oriented discussions. By categorizing key rhetorical relations the way we have, we have highlighted the dynamics of how content is grounded and how interactions evolve within negotiation spaces. Our findings suggest that opinion-discussion contexts foster denser negotiation spaces, where alignment and opposition actively shape the common ground, while fact-oriented dialogues emphasize clarification and mutual understanding.

We further show that biased questions are significantly more frequent in opinion-driven subreddits. This pattern reflects the dialogic nature of these communities, where questions are not merely requests for information but moves in a broader argumentative strategy, designed to provoke reflection, highlight implicit assumptions, or challenge interlocutors' positions.

Additionally, our analysis of intra-turn rhetorical structure reveals that while all communities make use of CONTINUATION to extend discourse, the presence of more structurally rich relations such as EXPLANATION/ELABORATION, CONTRASTIVE, and RESULT is indicative of greater discursive complexity. These relations are more prevalent in communities that favor elaborated and reasoned discourse, such as *askscience* and *changemyview*, whereas subreddits like *prochoice* show a more linear, monologic progression of thought.

Taken together, these findings contribute to a nuanced understanding of how rhetorical strategies vary with communicative context, offering a foundation for future work on discourse grammar and common ground modeling in online conversation.

References

Stergos Afantenos, Eric Kow, Nicholas Asher, and Jérémy Perret. 2015. Discourse parsing for multi-party chat dialogues. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 928–937, Lisbon, Portugal. Association for Computational Linguistics.

Jacopo Amidei, Paul Piwek, and Svetlana Stoyanchev. 2021. QTMM2012c+: A Queryable Empirically-grounded Resource of Dialogue with Argumentation.

Nicholas Asher, Farah Benamara, and Yvette Yannick Mathieu. 2009. Appraisal of Opinion Expressions in Discourse. *Lingvisticae Investigationes*, 32(2):279–292.

Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. Discourse Structure and Dialogue Acts in Multiparty Dialogue: the STAC Corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2721–2727, Portorož, Slovenia. European Language Resources Association (ELRA).

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Nicholas Asher and Alex Lascarides. 2008. Commitments, Beliefs and Intentions in Dialogue. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue (Londial)*, pages 35–42.

J. L. Austin. 1962. *How to Do Things with Words*. Oxford University Press.

Andrea Beltrama. 2018. Subjective assertions are weak: exploring the illocutionary profile of perspectivedependent predicates. *ZAS Papers in Linguistics*, 60:161–173.

Ellen Breitholtz. 2020. *Enthymemes and Topoi in Dialogue: The Use of Common Sense Reasoning in Conversation*. Brill. Accepted: 2021-04-22T15:02:05Z.

Herbert H. Clark. 1996. *Using Language*. 'Using' Linguistic Books. Cambridge University Press, Cambridge.

Herbert H. Clark and Edward F. Schaefer. 1989. Contributing to Discourse. *Cognitive Science*, 13(2):259–294.

Donka F. Farkas and Kim B. Bruce. 2010. On Reacting to Assertions and Polar Questions. *Journal of Semantics*, 27(1):81–118.

Bart Geurts. 2024. Common Ground in Pragmatics. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, winter 2024 edition. Metaphysics Research Lab, Stanford University.

Anastasia Giannakidou and Alda Mari. 2021a. Modalization and bias in questions. *University of Chicago and Insitut Jean Nicod*.

Anastasia Giannakidou and Alda Mari. 2021b. *Truth and Veridicality in Grammar and Thought: Mood, Modality, and Propositional Attitudes*. University of Chicago Press, Chicago, IL.

Jonathan Ginzburg. 1996. Dynamics and the semantics of dialogue. *Logic, language and computation*, 1:221–237.

Jonathan Ginzburg. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford University Press UK.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1):125–179. Place: Cambridge, MA Publisher: MIT Press.

Olusegun Oladele Jegede. 2025. Speech Acts in the Digital Sphere: A Corpus-Based Exploration of Interactional Dynamics on Social Media Platforms. *Corpus-based Studies across Humanities*, 2(2):359–383. Publisher: De Gruyter.

Shafiq Joty and Tasnim Mohiuddin. 2018. Modeling Speech Acts in Asynchronous Conversations: A Neural-CRF Approach. *Computational Linguistics*, 44(4):859–894.

Shafiq Rayhan Joty, Giuseppe Carenini, and Raymond T. Ng. 2013. Topic Segmentation and Labeling in Asynchronous Conversations. *Journal of Artificial Intelligence Research*, 47:521–573. ArXiv:1402.0586 [cs].

Manfred Krifka. 2015. Bias in commitment space semantics: Declarative questions, negated quetions, and question tags. *Semantics and Linguistic Theory*, 25:328.

Jonathan K. Kummerfeld, Sai R. Gouravajhala, Joseph J. Peper, Vignesh Athreya, Chulaka Gunasekara, Jatin Ganhotra, Siva Sankalp Patel, Lazaros C Polymenakos, and Walter Lasecki. 2019. A Large-Scale Corpus for Conversation Disentanglement. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3846–3856, Florence, Italy. Association for Computational Linguistics.

D. Ladd. 1981. A First Look at the Semantics and Pragmatics of Negative Questions and Tag Questions.

Pierre Larrivée and Alda Mari. 2022. Interpreting high negation in Negative Interrogatives: the role of the Other. *Linguistics Vanguard : a Multimodal Journal for the Language Sciences*, 8(s2):219–226. Publisher: De Gruyter.

Enzo Laurenti, Nils Bourgon, Farah Benamara, Alda Mari, Véronique Moriceau, and Camille Courgeon. 2022a. Give me your Intentions, I'll Predict our Actions: A Two-level Classification of Speech Acts for Crisis Management in Social Media. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4333–4343, Marseille, France. European Language Resources Association.

Enzo Laurenti, Bourgon Nils, Farah Benamara, Mari Alda, Véronique Moriceau, and Courgeon Camille. 2022b. Speech acts and Communicative Intentions for Urgency Detection. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 289–298, Seattle, Washington. Association for Computational Linguistics.

Stephen C. Levinson. 1983. *Pragmatics*. Cambridge [England] ; New York : Cambridge University Press.

David Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1):339–359.

Khyati Mahajan and Samira Shaikh. 2021. On the Need for Thoughtful Data Collection for Multi-Party Dialogue: A Survey of Available Corpora and Collection Methods. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 338–352, Singapore and Online. Association for Computational Linguistics.

Shikib Mehri and Giuseppe Carenini. 2017. Chat Disentanglement: Identifying Semantic Reply Relationships with Random Forests and Recurrent Neural Networks. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 615–623, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Andrei Popescu-Belis. 2007. Dialogue Acts: One or More Dimensions?

Paul Portner. 2009. *Modality*. Oxford Surveys in Semantics and Pragmatics. Oxford University Press, Oxford, New York.

Paul Portner. 2018. *Mood*. Oxford University Press.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *arXiv preprint*. ArXiv:2003.07082 [cs].

Liang Qiu, Yizhou Zhao, Weiyan Shi, Yuan Liang, Feng Shi, Tao Yuan, Zhou Yu, and Song-Chun Zhu. 2020. Structured Attention for Unsupervised Dialogue Structure Induction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1889–1899, Online. Association for Computational Linguistics.

Antonio Roque and David Traum. 2008. Degrees of Grounding Based on Evidence of Understanding. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 54–63, Columbus, Ohio. Association for Computational Linguistics.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1978. A Simplest Systematics for the Organization of Turn Taking for Conversation*. In Jim Schenkein, editor, *Studies in the Organization of Conversational Interaction*, pages 7–55. Academic Press.

Roser Saurí and James Pustejovsky. 2009. FactBank: A corpus annotated with event factuality. *Language Resources and Evaluation*, 43:227–268.

John R. Searle. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge.

Merel Semeijn. 2024. Common Ground in Non-face-to-face Communication: In Sensu Diviso or In Sensu Composito. *Journal of Philosophical Logic*, 53(3):657–678.

Robert Stalnaker. 1978. Assertion. *Syntax and Semantics (New York Academic Press)*, 9:315–332. Publisher: New York: New York Academic Press.

Radoslava Trnavac and Maite Taboada. 2012. The contribution of nonveridical rhetorical relations to evaluation in discourse. *Language Sciences*, 34(3):301–318.

Linzi Xing and Giuseppe Carenini. 2021. Improving Unsupervised Dialogue Topic Segmentation with Utterance-Pair Coherence Scoring. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 167–177, Singapore and Online. Association for Computational Linguistics.

Seth Yalcin. 2024. Defining common ground. *Linguistics and Philosophy*, 47(6):1045–1070.

A Appendix A: Examples of Speech Act Annotations

The following examples illustrate the speech act categories used in our annotation scheme. Each entry corresponds to a sentence or clause from the corpus labeled with one of the defined categories. In the case of ASSESSMENT and QUESTION (OPINION), we underline lexical markers of subjectivity (e.g., epistemic modals, evaluative adverbs).

ASSESSMENT

(1) You are definitely right that our human thinking can be our enemy here.

(2) That's probably the best outcome an English king could have hoped for.

ASSERTION

(3) Light is both an electric and magnetic wave.

(4) The United States does not recognize or consider Taiwan to be part of China.

QUESTION (OPINION)

(5) [*Straight:*] Would you say "both interpretations are valid" about other key biblical relationships or events?

(6) [*Biased:*] Wouldn't you rather have genuine public discourse shape foreign policy than the other way around?

QUESTION (FACT)

(7) [*Straight:*] What triggers a shark's blood sense/scent?

(8) [*Biased:*] Doesn't this require very pure water?

ADVICE/SUGGESTION/RECOMMENDATION

(9) Buy an extender from a big box store to move everything out to where you need it.

(10) Here is an example of someone selling one of the AM antennas on Etsy [URL], just to give you an idea of what they might look like.

EXPRESSIVE

(11) Haha okay yeah I get it now thanks.

(12) Apologies if my understanding is incorrect.

OTHER

(13) Painting Wood Posts

(14) FYI

B Appendix B: Examples of Rhetorical Relations Between Segments

The examples below illustrate rhetorical relations annotated in our corpus, distinguishing between those that occur *between different comments* (inter-comment) and those that occur *within the same comment* (intra-comment). Each turn is labeled using a speaker identifier (A, B, etc.) followed by a numerical index indicating the sequential order of the speech act within that speaker's contributions (e.g., **A.2** refers to the second speech act from Speaker A).

Inter-comment rhetorical relations. These involve rhetorical moves that span across distinct speakers or comment turns. The following examples illustrate common inter-comment rhetorical relations.

Agreement and Disagreement—
A.1: The yeast eat the priming sugar inside the bottle...
B.1: That's true for basic home brewing.
B.2: Larger and industrial operations instead carbonate the beer after the yeast has pretty much worked through all the sugars in the brew.

***B.1** expresses* AGREEMENT *with **A.1**; **B.2** contrasts with the generalization in **A.1**, expressing* DISAGREEMENT.

Request Clarification and Answer—
A.1: The mechanism is the same as you described either way...
B.1: Is it doing the exact same thing in beer, or is it different?
A.2: Exactly the same, yes.

B.1 REQUESTS CLARIFICATION *of **A.1**; **A.2** responds with a direct* ANSWER *to **B.1**.*

Continuation -EX- —
A.1: Buy an extender from a big box store to move everything...
B.1: Thank you!

***B.1** provides a minimal* CONTINUATION -EX- *follow-up to **A.1** across comments.*

Intra-comment rhetorical relations. These involve rhetorical relations internal to a single comment, typically reflecting internal discourse structure such as explanation, elaboration, contrast, continuation, or causal inference.

Explanation/Elaboration————————
A.1: My cat goes absolute apeshit for Subway's multigrain...
A.2: Turns out it's because there's tons of catnip oil...

***A.2** offers an* EXPLANATION/ELABORATION *for **A.1** within the same comment.*

Contrastive——————

A.1: (Older) Android smartphones use the headphone cable. . .

A.2: But they can only pick up FM signals.

***A.2** introduces a limitation that contrasts with the implication of **A.1**, establishing a* CONTRASTIVE *relation.*

Result——————

A.1: Many of them only have as much power as the people give them...

A.2: Which is why education and protest will be so important...

***A.2** presents a consequence of the proposition in **A.1**, marking a* RESULT *relation.*

Continuation——————

A.1: I just need to clarify a little bit of what you're saying...

A.2: You say they mean nothing,

***A.2** continues and specifies the content introduced in **A.1**, forming a* CONTINUATION *relation within the same comment.*