Degrees of Subjectivity and their Repercussions in Conversation
The View from Online Interactions

Gonzalo Freijedo Aduna[1], **Anastasia Giannakidou**[2], **Alda Mari**[1]
[1]Institut Jean Nicod, CNRS/ENS/EHESS/PSL University
[2]The University of Chicago
gonzalo.freijedo.aduna@ens.psl.eu, giannaki@uchicago.edu, alda.mari@ens.psl.eu

Abstract

We present the **Annotated Reddit Conversation Corpus (ARCC)**, an English-language dataset of online discussions annotated for *Speech Acts* and *Functional Dependence Relations*, designed to investigate how varying degrees of subjectivity influence conversational dynamics and interaction patterns. At the speech act level, we distinguish factual from opinion statements and further classify opinions along a five-degree scale of subjectivity. Functional Dependence Relations capture how segments relate to preceding ones. Analyses show that opinion-discussion contexts feature frequent inter-subjective opinions eliciting explicit agreement and disagreement, while information-exchange contexts exhibit less subjective opinions with responses like answers or requests for clarification. We further demonstrate that a transformer model can predict the subjectivity scale with promising performance. The corpus and annotation guidelines are made available to support future research on opinion expression and automated dialogue analysis.

Keywords: Pragmatic Annotation, Speech Acts, Functional Dependence Relations, Subjectivity

1. Introduction

Language is used not only to describe the world but also to express perspectives, evaluations, and normative stances (Portner, 2009). In everyday conversation, which today also includes online interaction, speakers constantly alternate between statements that aim to represent how the world objectively is and statements whose truth depends, wholly or partially, on an evaluation parameter. The latter are therefore subjective to varying degrees.

Our goal is to examine the pragmatic impact of subjectivity in online discourse, specifically how and to what extent subjective statements (classified along a graded subjectivity scale) influence the trajectory of interactions.

Formal models of dialogue have long emphasized that different conversational moves, particularly those carrying distinct illocutionary force, constrain the range of possible follow-up moves (Farkas and Bruce, 2010; Ginzburg, 2012; Krifka, 2022). On the other hand, research in NLP and computational linguistics has addressed subjectivity as a gradable phenomenon. However, to the best of our knowledge, the interaction between these two lines of inquiry (namely, how graded subjectivity affects the dynamics of conversational structure) has not yet been systematically investigated.

To address this, we introduce a corpus of Reddit conversations (**ARCC**, Freijedo Aduna et al., 2026) drawn from different Reddit communities[1]. Comments were segmented into minimal communicative units (or functional segments), each of which was annotated at two levels:

1. **Speech act annotations**, identifying the illocutionary force of each segment. At this level, we distinguish between statements and questions that provide or request factual information, and those that concern opinions. The latter were classified in a five-degree scale of subjectivity. This distinction enables us to differentiate between information-exchange contexts (primarily governed by the former) and opinion-discussion contexts.

2. **Functional Dependence Relation annotations**, capturing the communicative function a segment fulfills in relation to the segment to which it is linked. This layer of annotation allows us to examine whether, and how, the responses or follow-up contributions vary across different conversational contexts.

Our distinction between factual and opinion statements builds on a twofold distinction proposed by Giannakidou and Mari (2021a). The first dimension differentiates between *veridical* and *nonveridical* commitment, while the second contrasts *exogenous* and *endogenous* evidence. As justified in Section 2, we classify both nonveridical statements and statements grounded in endogenous evidence as opinion statements.

Extending this approach, we introduce a five-degree scale of subjectivity based on the type of evidence supporting the claim. The scale spans from maximally subjective judgments, such as predicates of personal taste, where the relevant evidence is personal, unshareable experience, to comparatively less subjective epistemic modals as (1),

[1] https://github.com/gFreijedo/arcc2026-subjectivity.

Proceedings of the Workshop on Structured Linguistic Data and Evaluation, pages 1–13
May 11, 2026.

whose truth conditions ultimately depend on publicly accessible evidence.

(1) Speaker (Seeing people arrive with wet umbrellas): It must be raining.

We experimented with a transformer model (`RoBERTa-base`) in order to test its capacity to learn the scale of subjectivity, obtaining relatively good results.

By operationalizing these distinctions through corpus annotation, we provide an empirical framework for investigating how degrees of subjectivity shape discourse structure and condition the types of reactions that emerge in different communicative contexts.

Our analyses show that opinion-discussion contexts are characterized by high frequencies of intersubjective opinions accompanied by explicit agreement and disagreement, whereas information-exchange contexts feature less subjective opinions and responses such as answers, continuers, or requests for clarification.

In this perspective, a natural downstream application of our work is the improved identification of opinion- versus fact-oriented statements in dialogue, taking into account both their degree of subjectivity and their interactional profile. In particular, given a sufficiently large corpus, it would be possible to model, for each segment, the distribution of responses it is likely to elicit (e.g., agreement, disagreement, elaboration, or clarification requests). Such predicted interactional patterns could then be used as additional features for distinguishing between fact- and opinion-oriented statements, as well as for refining the estimation of subjectivity degrees. More generally, this approach opens the way to integrating conversational dynamics into tasks such as stance detection, argument mining, and the analysis of deliberative or non-cooperative dialogue.

The paper is organized as follows. Section 2 presents our theoretically grounded binary distinction between fact and opinion statements, along with a finer-grained classification of opinions on a scale of subjectivity. Section 3 reviews related work. Section 4 describes the corpus collection and annotation procedure. Before presenting the annotation results in Section 6, we report a transformer-based evaluation of learnability in Section 5. Finally, Section 7 concludes the paper.

2. Conceptual definitions

In this section, we discuss the theoretical foundation of both the binary distinction between fact- and opinion-statements and the gradation of subjectivity that we propose within the latter.

	Evidence			
	Exogenous		Endogenous	
	Total	Partial	Total	Partial
Veridical commitment	Fact	----	Opinion	----
Nonveridical commitment	----	Opinion	----	Opinion

Table 1: Distinction between fact-statements and opinion-statements based on the *type* and the *amount* of evidence.

2.1. Fact- vs. Opinion-statements

For the binary distinction between fact-statements and opinion-statements, we rely on two notions issued from the semantic literature: *veridical commitment* (Giannakidou and Mari, 2021a,b) and *type* of evidence (Nuyts, 2001; Karttunen and Zaenen, 2005; Aikhenvald and Aikhenvald, 2006; Giannakidou and Mari, 2021a,b).

Taking a *veridical commitment* means that, by uttering a plain assertion such as (2) in a cooperative conversation, the speaker is fully committed to the truth of the proposition expressed by the utterance. In contrast, by using epistemically modalized sentences like (3), speakers signal a *nonveridical commitment*, which means that they leave open the possibility of the complement being false.

(2) Yeasts metabolize carbohydrates into carbon dioxide gas and ethanol (alcohol).[2]

(3) That may give you enough room to get the outlet out.

When following rationality and cooperative principles (Grice, 1975), speakers shape their contributions according to the evidence available. Thus, in (2), the speaker signals sufficient evidence on the topic, whereas in (3), the speaker explicitly indicates a lack thereof. Accordingly, we treat sentences like (3) as opinion-statements, in which the speaker's choice of words signals insufficient or partial evidence for their claims. That is, in our classification, ***nonveridical* statements are treated as opinion statements**. In this respect, we depart from Nuyts (2001), who argues that epistemic modals are not *necessarily* subjective.

However, this does not imply that all opinion-statements are *nonveridical*. On the contrary, speakers are, more often than not, fully committed to their opinions. To determine whether a fully committed speaker is asserting a fact or expressing an opinion, it is necessary to consider the *type* of the evidence on which their assertion can be based.

[2] All of the following examples are drawn from our corpus, except for those marked with an asterisk (*), which are constructed for illustrative purposes.

Following Giannakidou and Mari (2021a), we distinguish between **exogenous** and **endogenous** evidence. Exogenous evidence refers to informational content, that is, "a body of knowledge that includes 'facts'". In hearing (2), for instance, an interlocutor would assume that this kind of evidence is available. Endogenous evidence, by contrast, refers to emotive or affective content, including "desires and hopes [...], as well as political, ideological, religious, or aesthetic beliefs" (Giannakidou and Mari, 2021a). In uttering (4), the speaker is expressing an opinion based on his sociological or political beliefs.

(4) Abortion is Healthcare

We consider that, *alongside nonveridical statements*, **statements based on endogenous evidence are opinion-statements**. Figure 1 summarizes these considerations[3].

2.2. Degrees of subjectivity

However, as noted in the Introduction, expressions of opinion vary in their degree of subjectivity.

The epistemic modals discussed in the previous section are less subjective than, for instance, predicates of personal taste. The latter express highly subjective judgments, grounded solely in the speaker's personal experience. For this reason, the linguistic literature characterizes them as giving rise to cases of *faultless disagreement* (among numerous works, see MacFarlane 2005; Lasersohn 2005; Stephenson 2007; Sæbø 2009; Kennedy 2013; Umbach 2016; Coppock 2018).

We propose that expressions involving communal and moral values occupy an intermediate position. Although they are still based on endogenous evidence, that is, they reflect personal beliefs, they are intrinsically *inter-subjective*. Community and normative values are, by their very nature, oriented toward regulating social behavior and guiding interactions among individuals. In expressing opinions about such matters, speakers necessarily take others into account[4].

[3]The notion of partial endogenous evidence is more difficult to characterize, since speakers typically have full introspective access to their own beliefs. A more accurate formulation might be that of *conflicting* endogenous evidence, as in cases such as "Perhaps I should speak out, even if it's uncomfortable(*)". As this distinction is not central to our annotation scheme, we leave a more precise account for future work.

[4]We remain neutral with respect to the philosophical debate between moral realism and moral relativism, focusing instead on how moral evaluations function as intersubjectively recognized commitments within discourse, in line with John Searle's account of institutional reality (Searle, 1995).

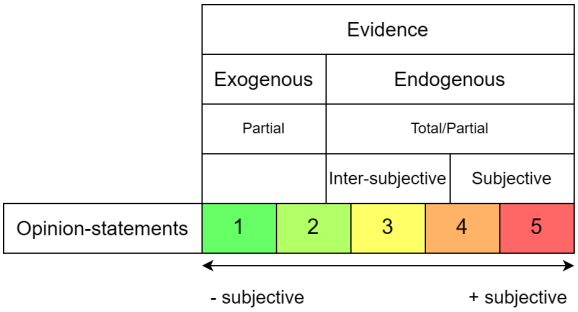

Table 2: Distinction between subjective and inter-subjective endogenous evidence and grades of subjectivity of opinion-statements.

So far, we have distinguished three levels of subjectivity, based on the type and amount of available evidence: the first corresponds to the facts of the world, or exogenous evidence when partial; the second to inter-subjective endogenous evidence; and the third to purely subjective endogenous evidence. However, as we will illustrate with examples from our corpus in Section 4, we also take into account cases in which multiple types of evidence are intertwined. In some instances, for example, an utterance simultaneously invokes personal taste and moral evaluation. Consider the following example:

(5) It would be fun if it weren't financed with dirty money*.

In this case, the speaker expresses a judgment of personal taste (it would be fun), while at the same time introducing a moral evaluation (financed with dirty money). The overall interpretation thus combines highly subjective experiential evidence with inter-subjective normative considerations.

There are also cases in which speakers evaluate real-world events on the basis of personal beliefs, thereby combining exogenous evidence with endogenous, inter-subjective evidence, as in (6). In such instances, reference is made to publicly accessible facts or events (in the example, historic events), but the assessment itself is filtered through normative or value-laden commitments. The resulting judgment is thus anchored in the external world while simultaneously reflecting the speaker's evaluative stance.

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

These considerations ultimately lead us to propose a five-degree scale of subjectivity (Figure 2). The scale ranges from (1-WORLD) statements with minimal subjectivity, whose truth can, in principle, be settled by facts of the world; (2-WORLD+VALUES) statements combining objective-world conditions and inter-subjective values; (3-VALUES) statements

based solely on inter-subjective or community values; (4-TASTE+VALUES) statements combining inter-subjective values and personal taste; and (5-TASTE) statements evaluated relative to an individual parameter, such as predicates of personal taste, representing the maximal level of subjectivity.

3. Related Work

In recent years, substantial effort within computational linguistics and NLP communities has been devoted to distinguishing factual from opinionated or subjective statements, often as a preparatory task for opinion mining and sentiment analysis. For a broad overview of frameworks, techniques, and applications in sentiment analysis, see Chandan and Mandal (2025). Although this factual–subjective distinction is frequently motivated by sentiment analysis tasks, our focus here is different: we concentrate on corpus annotation projects that model textual and conversational subjectivity, as well as recent work closely related to our own.

A seminal contribution to the study of subjectivity in natural language is Wiebe et al. (2005), who propose an annotation scheme for expressions of opinions and emotions at the phrase or clause level in news articles. Their framework provides a fine-grained representation of subjective content. Our work extends this line of research to conversational data, examining how expressions of opinion in Reddit comments shape and constrain subsequent contributions.

More recently, Antici et al. (2024) introduced a corpus of sentences extracted from news articles and classified as either subjective or objective, following highly prescriptive annotation guidelines. By contrast, Savinova and Moscoso Del Prado (2023) adopt the opposite approach, relying on naïve annotators who are provided only with brief definitions of objective and subjective statements. A key contribution of their work is to move beyond a purely binary distinction and to treat subjectivity as a gradable scale. In our study, we combine insights from both approaches: we provide theoretically informed annotation guidelines while modeling subjectivity expression as an ordinal five-level scale.

Another approach to the classification of subjective statements is proposed by Benamara et al. (2011), who introduce a taxonomy of elementary discourse units (EDUs) for the automatic detection of subjectivity in discourse (in this case, online film reviews). They distinguish four classes: (1) explicit subjectivity, identified using a subjectivity lexicon; (2) implicit subjectivity, whose interpretation must be resolved from context; (3) non-evaluative subjectivity, capturing propositional attitudes treated as separate EDUs from their complements; and (4) non-subjective EDUs. While the present study focuses on linguistically explicit expressions of subjectivity, we acknowledge the importance of accounting for their implicit realizations, which we leave for future work.

Closely related to our project, the influential FactBank corpus (Saurí and Pustejovsky, 2009) provides an annotation framework for capturing degrees of factuality in text, labeling events as true, possible, or counterfactual. Although FactBank, like the previously mentioned resources, is not based on conversational data, its emphasis on fine-grained factuality and on the interaction between modality and evidence informs our treatment of opinion statements and supports our distinction between factual and subjective contributions in dialogue.

Considering conversational corpora, a key annotation project is that of Stolcke et al. (2000), who focus on dialogue act modeling for the automatic tagging of conversational speech. Their classification includes a distinction between factual and opinion statements, grounded in pragmatic considerations. Building on this idea, we define finer-grained opinion categories to capture more nuanced conversational phenomena, informed by recent developments in semantic and pragmatic theory.

The STAC corpus (Asher et al., 2016) contains multi-party chat dialogues annotated for rhetorical relations and dialogue acts using SDRT (Asher and Lascarides, 2003). The ANNODIS project (Péry-Woodley et al., 2011) provides French online dialogues annotated for rhetorical relations and dialogue acts. Both serve as references for multi-level annotation of online conversations.

Similarly, the ISO standard 24617-2 (Bunt et al., 2012, 2020) offers a highly detailed framework for dialog act annotation. While this level of precision is valuable, it is often too complex for our purposes. We therefore adopt a simpler scheme, but borrow from ISO 24617-2 the notion of functional dependence relation to characterize how one segment relates to another.

Finally, a note on discourse parsing. A major challenge in the annotation process was reconstructing reply-to chains by linking each segment to the segment it responds to. Automatic discourse parsing has addressed related tasks through sequential modeling of discourse structure (Hernault et al., 2010), joint modeling of speech acts and discourse relations (Joty and Mohiuddin, 2018), and topic segmentation with relation labeling (Joty et al., 2013). In future work, we plan to explore similar approaches in order to enable the expansion of the corpus.

Subreddit	Posts	Comments	Segments
Askpolitics	3	40	145
DIY	8	78	254
askscience	11	111	448
changemyview	7	82	458
prochoice	5	99	315
todayilearned	13	159	310
Total	**47**	**569**	**1930**

Table 3: Corpus Statistics by Subreddit

4. Corpus and annotation procedure

In this section, we describe the collection of our corpus, the segmentation of the comments, the reorganization of the reply chains, and our two-level annotation scheme.

4.1. Corpus collection and segmentation

We collected the data by scraping six Reddit communities (subreddits): *AskPolitics*, *askscience*, *changemyview*, *DoItYourself*, *prochoice*, and *todayilearned*. These subreddits were selected to represent a range of communicative contexts.

Since Reddit comments often contain multiple communicative moves within a single turn, we segmented each comment into smaller units to annotate the parts of the dialogue that fulfill distinct communicative functions, called *functional segments* in ISO 24617-2 (Bunt et al., 2012). To achieve this, we first used *Stanza* (Qi et al., 2020) for tokenization, part-of-speech tagging, dependency parsing, and sentence segmentation. We then applied additional rule-based procedures to obtain finer-grained utterance units. This process consisted in detecting discourse markers (e.g., *but*, *because*, *although*) and multi-word connectors (e.g., *even though*, *so that*), while also relying on syntactic cues such as the presence of both subject and predicate to prevent the improper splitting of coordinated verb phrases. Each segment was assigned a unique identifier based on the original comment ID. Minor segmentation errors were subsequently corrected during the manual annotation phase. The resulting corpus comprises a total of 47 posts, 569 comments, and 1,930 segments; a breakdown by subreddit is provided in Table 3.

We subsequently restructured the reply-to chain so that each segment is explicitly connected to the precise segment it addresses. In the vast majority of cases, a segment is linked to the immediately preceding segment within the same comment. Two main exceptions are: (1) the first segment of a comment, which is typically linked to the final segment of the comment it replies to, and (2) segments that explicitly quote or directly address a non-adjacent segment. In (7), for instance, segment $B1$ is linked to $A1$ and $B2$, to $A2$.

(7) A: $\left[\text{What year did you graduate}\right]_{A1}$ $\left[\text{and how was it?}\right]_{A2}$
B: $\left[\text{2029,}\right]_{B1}$ $\left[\text{it was awesome}\right]_{B2}$

4.2. Annotation Scheme

The annotation was carried out at two levels: *Speech Acts* and *Functional Dependence Relations*.

By *Speech Acts*, we refer to the illocutionary force of an utterance in the sense of Searle (1969), i.e., the type of communicative act intended by the speaker when producing the utterance (e.g., asserting, questioning, advising). Each annotated speech act corresponds to a *functional segment* as defined in Bunt et al. (2020).

Functional Dependence Relations capture the communicative function of a segment with respect to a preceding segment. The notion is adopted from Bunt et al. (2020), who define it as a "relation between a given dialogue act and a preceding dialogue act on which the semantic content of the given dialogue act depends due to its communicative function". In this sense, Functional Dependence Relations are closely related to the linguistic notion of *Rhetorical Relations* in SDRT (Asher and Lascarides, 2003), as both describe structured dependencies between discourse units. However, Rhetorical Relations are more general, since they also account for coherence relations between segments within a discourse more broadly, not limited to dialogical interactions between speakers

To ensure the reliability of the annotation scheme, the first 500 segments were independently annotated by two PhD students. This resulted in an inter-annotator agreement of **Cohen's Kappa = 0.78**, indicating substantial agreement. The remaining 1,430 segments were annotated according to the established guidelines to complete the GOLD standard. The fine-grained Subjectivity Scale for opinion statements (ASEESSMENT in our annotation scheme) was annotated by a single annotator from the initial annotation team. Since it was not possible to compute inter-annotator agreement for this layer, we later evaluate the operationalizability of the scale through a supervised learning experiment (see Section 5 below), testing whether the distinctions can be reliably learned by a transformer-based model.

Speech Acts. We distinguish the following speech act categories and refer the reader to Appendix A for illustrative examples drawn from our corpus:

- ASSESSMENT (ASSESS): statements describing how the world is or should be from the speaker's perspective.

- QUESTION (OPINION) (Q_OPIN): questions seeking an assessment as a response.
- ASSERTION (ASSERT): statements presenting facts without linguistically marked subjectivity.
- QUESTION (FACT) (Q_FACT): questions seeking factual information[5].
- EXPRESSIVE (EXPR): expressions of emotion, gratitude, greetings, irony, or humor.
- ADVISE (ADV): utterances intended to guide or influence the hearer's actions (typically, though not necessarily, in imperative form).
- OTHER (OTH): utterances that do not fall into the previous categories.

Subjectivity Scale for Assessments. ASSESSMENT instances were further subdivided into five subclasses reflecting increasing degrees of subjectivity based on the considerations presented in Section 2.2. In the examples below, the relevant linguistic cues are underlined:

- WORLD (WOR): opinions about factual matters (e.g., "That may give you enough room to get the outlet out").
- WORLD + VALUES (V/W): opinions about factual matters involving evaluation based on communal or inter-subjective values (e.g., "Putin's Russia is a mafia state").
- VALUES (VAL): opinions grounded in communal values, worldviews, or moral standpoints (e.g., "drones should be banned for civilian use").
- TASTE + VALUES (T/V): opinions based on personal taste intertwined with communal values (e.g., "but cool idea nonetheless").
- TASTE (TST): opinions grounded purely in personal taste predicates (e.g., "They still look as good as the day they went outside.").

Functional Dependence Relations. We annotate the relation between each segment and the segment it is linked to using the following categories, and refer the reader to Appendix B for illustrative examples from our corpus:

- AGREEMENT (AGR): explicit agreement with the linked segment.
- DISAGREEMENT (DIS): explicit disagreement with the linked segment.
- ANSWER (ANS): a statement responding to a linked question.
- REQUEST FOR CLARIFICATION (REQ): a question or statement seeking further specification of the linked segment.
- CONTINUATION–IN (CONT_IN): a segment linked to the immediately preceding segment within the same comment.
- CONTINUATION–EX (CONT_EX): a segment linked to a segment in another comment without signaling explicit agreement or disagreement.
- OFF-TOPIC (OFF): a segment that diverges from the topic of the linked segment.

[5]Both question types were additionally marked for *bias*, since interlocutors may agree or disagree with the presupposed or implied stance conveyed by a question.

5. Model-Based Evaluation of the Scale

To assess the internal coherence and operationalizability of our five-degree subjectivity scale, we first extracted all ASSESSMENT segments from the corpus. These opinion-bearing segments were then used as input to a supervised learning experiment with `RoBERTa-base` (Liu et al., 2019), following the general approach of Savinova and Moscoso Del Prado (2023). The five labels (WOR, V/W, VAL, T/V, TST) were mapped to numerical values from 1 to 5 and then normalized to the range 0–1 for regression. The model was trained with mean squared error loss, a small learning rate (1e-5), and early stopping.

Given the relatively small size of our dataset (835 English segments), we perform stratified 5-fold cross-validation to ensure stable and reliable performance. The model shows consistent results across folds, with *Mean Absolute Error* (MAE) = 0.1568 $\pm$ 0.0108 and *Root Mean Squared Error* (RMSE) = 0.2104 $\pm$ 0.0134, along with moderate-to-strong correlations (*Pearson's* r = 0.6689 $\pm$ 0.0475, *Spearman's* ρ = 0.6379 $\pm$ 0.0493), indicating stable performance across splits.

We then train a final model on an 80% training split and evaluate it on a held-out 20% test set. On the held-out data, the model achieves MAE = 0.1000, RMSE = 0.1465, Pearson r = 0.8563, and Spearman ρ = 0.8298, indicating strong linear and ordinal correspondence with the gold labels.

Figure 1 shows the linear regression between gold and predicted values on the test set ($R^2 = 0.733$), indicating that the model's predictions closely track the annotated subjectivity scores. The plot also reveals a slight bias at the extremes: for the least subjective category (1), the model predicts higher subjectivity than gold, while for the most subjective category (5), it predicts lower subjectivity.

Figure 2 presents the confusion matrix after rounding predictions to the nearest category, with 74.4% exact accuracy. Most misclassifications occur between adjacent categories, while confusions across distant categories are rare, supporting the

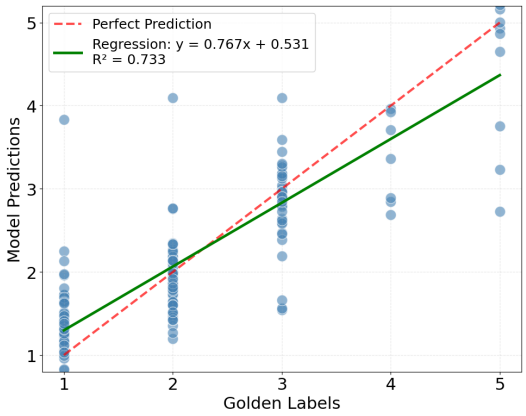

Figure 1: Linear regression between gold and predicted subjectivity scores on the held-out test set ($R^2 = 0.733$).

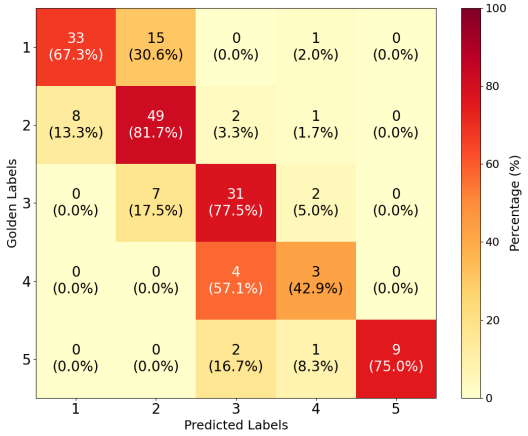

Figure 2: Confusion matrix showing exact-match prediction accuracy of 74.4%, computed across all samples and weighted by class frequencies.

graded structure of the scale. The most notable misclassification is for gold label 4 (T/V), which the model predicts as 3 in 57% of cases, showing occasional underestimation of higher degrees of subjectivity.

This pattern is partly due to the small number of examples for the high-subjectivity categories in our corpus (T/V and TST), which makes predictions for these classes less reliable. Additional annotated data for these categories would be necessary to achieve more conclusive results across the full scale of subjectivity.

We manually inspected the model's predictions that deviate by more than one degree from the gold label in order to better understand systematic sources of error. Due to space constraints, we illustrate this analysis with a small number of representative examples.

In (8), the model appears to assign greater weight to the expression "it's odd", interpreting it as an evaluation grounded in personal taste. However, under our annotation scheme, the property being evaluated (the absence of information about the person) is a fact about the world. Therefore, the correct label is 2, as the judgment evaluates an external, world-related state of affairs.

(8) Its odd but there's practically nothing about this guy in a Google Search. (True label: 2 — Values + World; Predicted label: 4 — Taste + Values)

In (9), the model seems to focus on the category "American", potentially interpreting the statement as invoking socially shared values or norms. However, the relevant evaluative component is "feel fake", which in our framework constitutes fully endogenous evidence. As such, the correct label is 5, reflecting a judgment grounded in personal taste rather than externally anchored values.

(9) And most of the "American" profiles feel fake. (True label: 5 — Taste; Predicted label: 3 — Values)

Although missing the correct label by only one degree, (10) highlights an important limitation of our approach: the model does not incorporate conversational context, in particular the content of the segment being responded to. In such cases, the interpretation of the utterance depends entirely on its antecedent, and the appropriate label is inherited from the preceding segment. The model's inability to account for this dependency leads to systematic misclassification of context-dependent responses.

(10) Exactly! (True label: 3 — Values; Predicted label: 4 — Taste + Values)

It is important to note that this experiment does not replace inter-annotator agreement; rather, it provides computational evidence that the annotation scheme encodes a coherent and learnable dimension of subjectivity grounded in systematic linguistic cues. It is also worth noting that the results reported correspond to the best-performing run, as performance shows non-negligible variance across different data splits. Future work could address this instability through data augmentation or ensemble techniques. The best-trained model is available in our repository.

6. Results

In this section, we present the main findings of our annotation, considering both the binary distinction between statements of fact and opinion, as well as the finer-grained degrees of subjectivity.

Subreddit	ASSESS	Q_OPIN	ASSERT	Q_FACT	EXPR	ADV	OTH	TOTAL
AskPolitics	84 (57.9%)	14 (9.7%)	38 (26.2%)	0 (0.0%)	4 (2.8%)	4 (2.8%)	1 (0.7%)	145
changemyview	343 (74.9%)	22 (4.8%)	63 (13.8%)	6 (1.3%)	12 (2.6%)	6 (1.3%)	6 (1.3%)	458
prochoice	164 (52.1%)	26 (8.3%)	56 (17.8%)	3 (1.0%)	51 (16.2%)	10 (3.2%)	5 (1.6%)	315
askscience	84 (18.8%)	4 (0.9%)	278 (62.1%)	61 (13.6%)	11 (2.5%)	10 (2.2%)	0 (0.0%)	448
DoItYourself	65 (25.6%)	4 (1.6%)	84 (33.1%)	26 (10.2%)	25 (9.8%)	46 (18.1%)	4 (1.6%)	254
todayilearned	96 (31.0%)	8 (2.6%)	144 (46.5%)	22 (7.1%)	28 (9.0%)	8 (2.6%)	4 (1.3%)	310
TOTAL	**836**	**78**	**663**	**118**	**131**	**84**	**20**	**1930**

Table 4: Distribution of speech acts across subreddits (counts with row-wise percentages).

6.1. Binary Classification

We first examined the distribution of speech acts across the six subreddits in our corpus (Table 4). For analysis, we grouped the subreddits into two *Global Communicative Contexts* based on the expected type of interaction: *opinion-discussion* (*AskPolitics*, *changemyview*, *prochoice*) and *information-exchange* (*askscience*, *DoItYourself*, *todayilearned*). Figure 3 presents a heatmap of Pearson residuals, highlighting the under- and over-representation of each speech act category within each group relative to the expected distribution under independence. In Opinion-Discussion contexts, ASSESSMENT and QUESTION (OPINION) are over-represented, while ASSERTION and QUESTION (FACT) are under-represented. The opposite pattern is observed in Information-Exchange contexts, reflecting their focus on factual information rather than evaluative or opinion-based contributions.

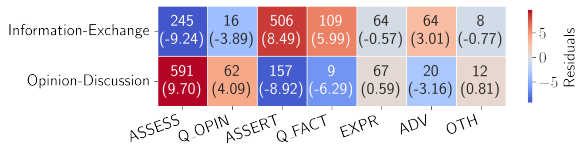

Figure 3: Heatmap of Pearson residuals showing the over- and under-representation of speech act categories across the two Global Communicative Contexts.

To examine how the two global communicative contexts shape conversational structure and dynamics, we then analyzed the distribution of Functional Dependence Relations in each context, shown in Figure 4. The results reveal distinct rhetorical profiles aligned with each context's communicative function.

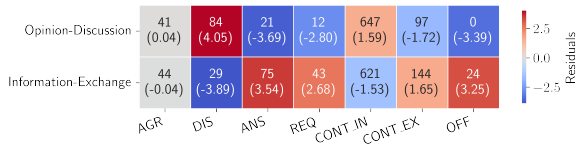

Figure 4: Heatmap of Pearson residuals for Functional Dependence Relations by global communicative context.

Opinion-discussion contexts show a strong over-representation of DISAGREEMENT (DIS) relations, reflecting the deliberative and argumentative nature of these communities where debate and negotiation of opposing viewpoints are central. In contrast, information-exchange contexts exhibit over-representation of ANSWER (ANS) and REQUEST FOR CLARIFICATION (REQ) relations, consistent with communities focused on collaborative knowledge sharing and uncertainty resolution.

We can also analyze patterns of Functional Dependence Relations by focusing on the immediate, local context, examining the types of relations that segments labeled as ASSERTION and ASSESSMENT receive from segments in replying comments (i.e., from other speakers) without considering the broader communicative context. As shown in Figure 5, ASSESSMENT segments are more likely to receive responses that explicitly agree or disagree with their content, whereas ASSERTION segments tend to elicit requests for clarification or continuers, consistent with the trends observed in the global discourse context.

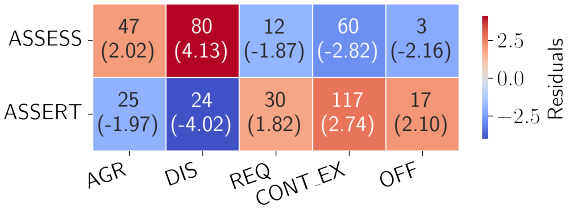

Figure 5: Heatmap of Pearson residuals for Functional Dependence Relations by local communicative context.

6.2. Five-Grade Subjectivity Classification

Moving beyond the binary fact/opinion distinction, we examine the distribution of opinions across our fine-grained five-grade subjectivity scale. This scale captures gradations of subjectivity, from opinions based on personal taste (TST) to opinions about world facts (WOR), with taste+value (T/V), values (VAL), and world facts+values (V/W) occupying intermediate positions.

Table 5 shows raw count and percentages in the whole corpus. Figure 6 shows the distribution of

SUB	Count	Proportion (%)
WOR	245	29.31
V/W	297	35.53
VAL	200	23.92
T/V	37	4.43
TST	57	6.82

Table 5: Distribution of Sub-Assessment Labels in the Corpus

these five opinion types across two global communicative contexts identified earlier.

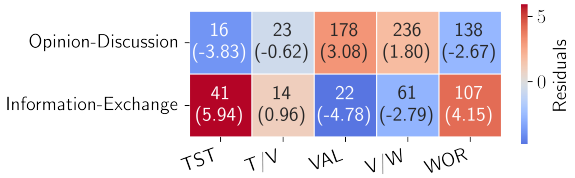

Figure 6: Heatmap of Pearson residuals for Sub-classes of ASSESSMENT by global communicative context.

In Information-Exchange contexts (*askscience*, *DoItYourself*, *todayilearned*), we observe over-representation of the most subjective category (TST) and, to a lesser extent, taste+value-based statements (T/V), alongside the most objective category (WOR). This pattern suggests that factual discussions occasionally incorporate both expressions of personal preference and claims about world states that approach fact-adjacent statements.

In contrast, Opinion-Discussion contexts (*AskPolitics*, *changemyview*, *prochoice*) show over-representation of value-based statements (VAL) and word+values statements (V/W), that is, opinion types that blend subjective judgment with appeals to shared values or factual claims.

Taken together with the patterns of Functional Dependence Relations observed in each Global Communicative Context in the previous Section, these results indicate that opinion statements grounded in intersubjective values are more likely to be explicitly discussed, eliciting direct expressions of agreement and disagreement.

However, given the relatively low frequency of the most subjective statements and the specific nature of the conversations in our corpus, these findings should be interpreted cautiously. They are not fully generalizable but can serve as preliminary indicators and guideposts for future research on subjectivity and interaction in online discourse.

7. Conclusions

We presented a corpus of Reddit conversations annotated for speech acts and Functional Dependence Relations. At the first level of annotation, we distinguished between fact and opinion statements, taking into account the amount and type of available evidence, and further characterized opinion statements along a five-degree scale of subjectivity.

Our analysis revealed systematic differences across communicative contexts. Opinion-Discussion contexts (*AskPolitics*, *changemyview*, *prochoice*) are characterized by a high frequency of opinion statements based on inter-subjective values, coupled with prominent Functional Dependence Relations of agreement and disagreement. In contrast, Information-Exchange contexts (*askscience*, *DoItYourself*, *todayilearned*) feature opinionated statements of lower subjectivity, with Functional Dependence Relations such as answers, continuers, and requests for clarification. These patterns highlight how discourse structure and interaction strategies are shaped by the type of communicative context.

Importantly, we fine-tuned a transformer model on the subjectivity scale, obtaining promising results, which opens the possibility for automatic or semi-automatic annotation. Such methods would allow us to extend the corpus and derive more generalizable conclusions. Future work also aims to annotate Functional Dependence Relations between segments within the same comment, capturing internal contribution structures, and further characterizing dialogue dynamics across different contexts.

8. Ethical Considerations and Limitations

The collection and annotation of Reddit comments prioritize user privacy by anonymizing all usernames and excluding original text. Users did not explicitly provide consent, as the data consists of publicly available content. The corpus has limitations in terms of generalizability: it includes only 47 conversations from a small set of specific subreddits, so findings may not extend to other Reddit communities or online discussions more broadly. Additionally, the relatively small size of the dataset constrains the representativeness of patterns observed. Finally, our study is restricted to English data, which may obscure cross-cultural variation in the interpretation of subjectivity. In particular, the assignment of categories such as VALUES or even the characterization of WORLD-related knowledge may depend on cultural and contextual factors. As such, the proposed annotation scheme may not transfer directly to multilingual or multicultural settings without further validation[6].

[6]We thank a reviewer for highlighting this issue. We also thank all reviewers, whose recommendations have substantially improved the quality of this paper.

9. Availability

The full corpus, along with the trained model of Section 5, is available at https://github.com/gFreijedo/arcc2026-subjectivity . To comply with Reddit's Terms of Service and GDPR regulations, ARCC does not redistribute raw comment text or original user identifiers. Usernames have been replaced with non-reversible, dataset-internal speaker IDs. The textual content of each segment can be reconstructed via the Reddit API using the provided comment identifiers and character offsets. The repository includes code for this reconstruction process.

It should be noted that some comments may no longer be accessible through the Reddit API due to deletion by users or moderators, which may prevent full reconstruction in certain cases.

10. Bibliological references

Alexandra Y. Aikhenvald and Alexandra Y. Aikhenvald. 2006. *Evidentiality*. Oxford University Press, Oxford, New York.

Francesco Antici, Federico Ruggeri, Andrea Galassi, Katerina Korre, Arianna Muti, Alessandra Bardi, Alice Fedotova, and Alberto Barrón-Cedeño. 2024. A Corpus for Sentence-Level Subjectivity Detection on English News Articles. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 273–285, Torino, Italia. ELRA and ICCL.

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.

Farah Benamara, Baptiste Chardon, Yannick Mathieu, and Vladimir Popescu. 2011. Towards Context-Based Subjectivity Analysis. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 1180–1188, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Harry Bunt, Jan Alexandersson, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Volha Petukhova, Andrei Popescu-Belis, and David Traum. 2012. ISO 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 430–437, Istanbul, Turkey. European Language Resources Association (ELRA).

Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prévot. 2020. The ISO Standard for Dialogue Act Annotation, Second Edition. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 549–558, Marseille, France. European Language Resources Association.

Manish Kumar Chandan and Shrabanti Mandal. 2025. A comprehensive survey on sentiment analysis: Framework, techniques, and applications. *Computer Science Review*, 58:100777.

Elizabeth Coppock. 2018. Outlook-Based Semantics. *Linguistics and Philosophy*, 41(2):125–164.

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

Anastasia Giannakidou and Alda Mari. 2021a. A Linguistic Framework for Knowledge, Belief, and Veridicality Judgment. *KNOW: A Journal on the Formation of Knowledge*, 5(2):255–293.

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

H. P. Grice. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics: Vol. 3: Speech acts*, pages 41–58. Academic Press, New York.

Hugo Hernault, Danushka Bollegala, and Mitsuru Ishizuka. 2010. A Sequential Model for Discourse Segmentation. In *Computational Linguistics and Intelligent Text Processing*, pages 315–326, Berlin, Heidelberg. Springer.

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].

Lauri Karttunen and Annie Zaenen. 2005. Veridicity. In *Annotating, Extracting and Reasoning about Time and Events*, volume 5151 of *Dagstuhl Seminar Proceedings (DagSemProc)*, pages 1–9, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Christopher Kennedy. 2013. Two Sources of Subjectivity: Qualitative Assessment and Dimensional Uncertainty. *Inquiry*, 56.

Manfred Krifka. 2022. Adjacency Pairs in Common Ground Update: Assertions, Questions, Greetings, Offers, Commands. In *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*, Dublin, Ireland. SEMDIAL.

Peter Lasersohn. 2005. Context Dependence, Disagreement, and Predicates of Personal Taste*. *Linguistics and Philosophy*, 28(6):643–686. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 6.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv:1907.11692 [cs].

John MacFarlane. 2005. The Assessment Sensitivity of Knowledge Attributions. In Tamar Szabó Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology*, pages 197–234. Oxford University Press.

Jan Nuyts. 2001. Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics*, 33(3):383–400.

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

Marie-Paule Péry-Woodley, Stergos D. Afantenos, Lydia-Mai Ho-Dac, and Nicholas Asher. 2011. Le corpus ANNODIS, un corpus enrichi d'annotations discursives [The ANNODIS corpus, a corpus enriched with discourse annotations]. *Traitement Automatique des Langues*, 52(3):71–101. Place: France.

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

Elena Savinova and Fermin Moscoso Del Prado. 2023. Analyzing Subjectivity Using a Transformer-Based Regressor Trained on Naïve Speakers' Judgements. In *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 305–314, Toronto, Canada. Association for Computational Linguistics.

John Searle. 1995. *The Construction of Social Reality*. Free Press.

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

Tamina Stephenson. 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4):487–525. Company: Springer Distributor: Springer Institution: Springer Label: Springer Number: 4.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–374. Place: Cambridge, MA.

Kjell Johan Sæbø. 2009. Judgment ascriptions. *Linguistics and Philosophy*, 32(4):327–352.

Carla Umbach. 2016. Evaluative propositions and subjective judgments. In Cécile Meier and Janneke Van Wijnbergen-Huitink, editors, *Subjective Meaning*, pages 127–168. De Gruyter.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2):165–210.

11. Language Resource References

Freijedo Aduna, Gonzalo and Giannakidou, Anastasia and Mari, Alda. 2026. *Annotated Reddit Conversation Corpus (ARCC) for Speech Acts, Functional Dependence Relations, and Subjectivity*. ISLRN 00000000.

Qi, Peng and Zhang, Yuhao and Zhang, Yuhui and Bolton, Jason and Manning, Christopher D. 2020. *Stanza: A Python NLP Library for Many Human Languages*. PID https://stanfordnlp.github.io/stanza/.

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

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

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

ASSERTION

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

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

QUESTION (OPINION)

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

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

QUESTION (FACT)

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

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

ADVICE/SUGGESTION/RECOMMENDATION

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

(20) 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

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

(22) Apologies if my understanding is incorrect.

OTHER

(23) Painting Wood Posts

(24) FYI

B. Appendix B: Examples of Functional Dependence Relations Between Segments

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

AGREEMENT AND DISAGREEMENT

(25) a. **A.1:** The yeast eat the priming sugar inside the bottle...
b. **B.1:** That's true for basic home brewing.
c. **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

(26) a. **A.1:** The mechanism is the same as you described either way...
b. **B.1:** Is it doing the exact same thing in beer, or is it different?
c. **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-

(27) a. **A.1:** Buy an extender from a big box store to move everything...
b. **B.1:** Thank you!

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

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

EXPLANATION

(28) a. **A.1:** My cat goes absolute apeshit for Subway's multigrain...
b. **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

(29) a. **A.1:** (Older) Android smartphones use the headphone cable. . .
b. **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

(30) a. **A.1:** Many of them only have as much power as the people give them...
b. **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

(31) a. **A.1:** I just need to clarify a little bit of what you're saying...
b. **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.*