

Annotated Reddit Conversation Corpus (ARCC)

Annotation Guidelines

February 2026

1 Introduction

The annotation aims to examine how varying degrees of subjectivity influence conversational dynamics in online discussions. Subjectivity is treated as a gradient reflecting the speaker’s evaluative stance and epistemic commitment. By annotating speech acts and functional relations, we investigate how different levels of subjectivity shape interactional patterns such as agreement, disagreement, continuator, and conversational development in general.

The annotation scheme operates at two levels: **Speech Acts** and **Functional Dependence Relations**.

At the level of speech acts, utterances are characterized in terms of communicative intention and illocutionary force. A primary distinction is drawn between **ASSERTION** (fact-oriented statements) and **ASSESSMENT** (opinion-oriented statements). **ASSESSMENTS** are further subcategorised along a five-degree scale of subjectivity, capturing gradience in evaluativity and degrees of epistemic or normative commitment.

At the level of **Functional Dependence Relations**, the scheme captures structural and argumentative dependencies between discourse units, specifying how segments relate to and depend on one another within the conversational exchange.

1.1 General Principles

Annotations are primarily based on **surface form**. In most cases, the grammatical form of a segment provides sufficient information to determine its speech act. For instance, interrogative syntax reliably signals questions, while declarative syntax typically corresponds to assertions or assessments. Likewise, task-oriented imperatives (e.g., instructions or recommendations directed at another participant) are systematically annotated as **ADVISE**.

However, surface form is not always sufficient. Some segments consist of **subsentential units** (e.g., short answers or discourse markers such as “yes,” “no,” “exactly,” “right”), whose interpretation depends on the preceding discourse. In such cases, annotators must consult the **parent segment** (i.e., the segment to which the current unit directly responds).

The speech act label of these minimal responses is determined by the speech act of the parent segment. For example:

- A response such as “yes” to an **ASSERTION** is annotated as an **ASSERTION**.
- A response such as “yes” to an **ASSESSMENT** is annotated as an **ASSESSMENT**.

In other words, when surface form does not provide sufficient information, the annotation should rely on the functional relation to the parent segment.

A more detailed account is provided in Sections 3 and 4.

1.2 Segmentation

The unit of annotation is the *functional segment* as defined in Bunt et al. [2020]. Annotators receive comments already segmented into functional units.

The reply-to chain is reconstructed and delivered in its finalized form, so that each segment is already linked to the segment it directly responds to. Annotators therefore do not need to determine conversational links themselves, but can rely on the provided discourse structure when assigning speech act and relational labels.

2 Two-level Annotation Scheme

Category	Subcategory	Label
ASSERTION	—	ASSERT
ASSESSMENT	WORLD	WOR
	WORLD + VALUES	V/W
	VALUES	VAL
	TASTE + VALUES	T/V
	TASTE	TST
QUESTION (FACT)	—	Q_FACT
QUESTION (OPINION)	—	Q_OPIN
ADVISE	—	ADV
EXPRESSIVE	—	EXPR
OTHER	—	OTH

Relation	Label
AGREEMENT	AGR
DISAGREEMENT	DIS
ANSWER	ANS
REQUEST FOR CLARIFICATION	REQ
CONTINUATION-IN	CONT_IN
CONTINUATION-EX	CONT_EX
OFF-TOPIC	OFF

Table 2: Functional Dependence Relation labels.

Table 1: Speech Act labels.

3 Speech Act Annotation

3.1 Assessment (ASSESS)

Definition: Statements presenting linguistically marked subjectivity. They describe how the world is or should be from the speaker’s perspective, rather than presenting information as objectively verifiable. Such statements express evaluation, judgment, uncertainty, likelihood, desirability, or personal stance.

Linguistic markers: Predicates of personal taste (e.g., *fun*, *tasty*, *interesting*, *boring*, *looks good*); modal auxiliaries (both in their deontic and epistemic flavors: e.g., *might*, *may*, *could*, *should*, *must*); modal adjectives and adverbs (e.g., *probable*, *probably*, *likely*, *possibly*, *perhaps*); vague or gradable adjectives and quantifiers (e.g., *many*, *few*, *several*, *a lot of*, *too much*, *significant*); evaluative adjectives (e.g., *good*, *bad*, *important*, *ridiculous*, *unfair*); evaluative nouns (e.g., *success*, *failure*, *mistake*, *disaster*, *achievement*, *problem*); and non-factive propositional attitude verbs (e.g., *believe*, *think*, *suppose*, *guess*, *assume*, *suspect*, *seem*).

3.1.1 World (WOR)

Definition: opinions about factual matters. Typically, epistemically modalized sentences indicating lack of full commitment. We include in this class vague predicates about factual matters.

Examples:

- (1) That may give you enough room to get the outlet out.
- (2) Lots of info available with some quick searches though...
- (3) it has a history of also having a certain structure to its governance
- (4) Some foods are a lot easier to digest than others.

3.1.2 World+Values (V/W)

Definition: Opinions about factual matters involving evaluation based on communal or inter-subjective values. We use ‘values’ in a broad sense: not only moral or explicitly normative standards, but also socially shared criteria such as economic worth, political legitimacy, social desirability, effectiveness, or success. These statements typically describe the world while simultaneously evaluating it, often through value-laden nouns or adjectives.

Examples:

- (5) Putin’s Russia is a mafia state.
- (6) Cars are expensive as hell period.
- (7) That’s probably the best outcome an English king could have hoped for.

- (8) At least Missouri has legal abortion

3.1.3 Values (VAL)

Definition: Opinions grounded primarily in communal values, ideological commitments, or moral standpoints, rather than in the description of a particular factual situation. These statements typically express what ought to be the case, what is right or wrong, acceptable or unacceptable, or how abstract social categories should be understood. They are frequently deontically modalized (e.g., *should*, *must*, *ought to*), but may also take the form of categorical value claims about norms, rights, or moral classifications.

Examples:

- (9) Drones should be banned for civilian use.
(10) Abortion is healthcare.
(11) I would do the same if one of my renters did it to any of my houses.

3.1.4 Taste+Values (T/V)

Definition: Opinions based on personal taste intertwined with communal or inter-subjective values. These statements combine individual experiential evaluation (e.g., aesthetic judgment or personal preference) with reference to socially shared standards, norms, or more generalizable assessments.

Examples:

- (12) But cool idea nonetheless.
(13) This site used to feel very America-centric when I first joined.
(14) Just because you personally don't enjoy them doesn't mean they universally suck.

3.1.5 Taste (TST)

Definition: Opinions grounded purely in personal taste predicates. These statements express individual, experiential, or aesthetic evaluations that depend on the speaker's subjective preferences, rather than on communal norms, moral principles, or broader ideological commitments. They typically involve adjectives or predicates of personal taste and are not presented as claims about what ought to be the case.

Examples:

- (15) They still look as good as the day they went outside.
(16) If it's the same vid I watched, the flooring choice was terrible.
(17) Very nice home in Moon Valley.

3.2 Assertion (ASSERT)

Definition: Statements presenting facts without linguistically marked subjectivity. Statements reporting third parties opinions are considered ASSERTION.

- (18) Then a storm built a land bridge.
(19) I did this with PT wood to build planters a few years ago.
(20) How that manifests in its constitution in particular I don't know.
(21) No Jews believe the rest of the Torah [...] are the literal word of God.

3.3 Question (Opinion) (Q_OPIN)

Definition: Questions that seek an ASSESSMENT as a response, or that themselves contain evaluative, value-laden, or stance-marking expressions. These questions typically invite a judgment, interpretation, or normative evaluation rather than purely factual information.

- (22) Why would someone laugh at this?
- (23) What happened to the BS line of AbOrTiOn ShOuLd Be LeFt To ThE sTaTeS?
- (24) Is this set up unsafe?

3.4 Question (Fact) (Q_FACT)

Definition: Questions that seek factual, descriptive, or explanatory information about states of affairs, events, or causal mechanisms. They request information that can, in principle, be verified independently of the speaker's evaluative stance.

- (25) How do you secure it at the edges?
- (26) What causes 'steam' over a cold body of water?
- (27) When was the policy implemented?

3.5 Advise (ADV)

Definition: Utterances intended to guide, recommend, or influence the hearer's actions (typically, though not necessarily, in imperative or suggestive form). They usually take the linguistic form of what we label as ASSESSMENT, but their primary function is 'task-oriented': to prompt, direct, or shape behavior rather than merely to express an opinion. The sharing of a link its considered an ADVICE (in the sense of a suggestion).

Examples:

- (28) Go look at askreddit posts;
- (29) Don't obey in advance.
- (30) I'd spruce it up a little more with a handrail.
- (31) [Here is an example of someone selling one of the AM antennas on Etsy](URL).

3.6 Expressive (EXPR)

Definition: Expressions of emotion, gratitude, greetings, irony, or humor. These utterances often convey the speaker's affective state rather than propositional content. They are frequently marked by typographical or paralinguistic cues such as exclamation marks, repeated punctuation, interjections, laughter indicators (e.g., "lol," "haha"), or emotive punctuation.

Examples:

- (32) thank you!
- (33) Lol
- (34) Please help!!!
- (35) I have felt so damn uneasy the last couple weeks.

3.7 Other (OTH)

Definition: Utterances that do not fall into the previous categories.

- (36) Here have a delta
- (37) FYI

4 Functional Dependence Relation Annotation

4.1 Agreement (AGR)

Definition: Explicit agreement with the linked segment. Agreement can be marked with lexical cues (e.g., “exactly,” “I agree,” “of course”) or through subtle phrasing that endorses or confirms the content of the previous segment. Short responses or elaborations that confirm the parent segment also count as AGREEMENT.

Examples:

- (38) a. The yeast eat the priming sugar inside the bottle to create carbonation
 b. That's true for basic home brewing,
- (39) a. cmv: both interpretations of David and johnathan are valid
 b. Well both are very possible.
- (40) a. this site used to feel very america-centric when I first joined.
 b. Yes, Reddit demographics has gotten more international and diverse in its views...

4.2 Disagreement (DIS)

Definition: Explicit disagreement with the linked segment. Disagreement can be expressed directly (e.g., “I disagree,” “not true”) or indirectly, by contradicting, challenging, or providing an alternative to the claim made in the parent segment.

Examples:

- (41) a. The yeast eat the priming sugar inside the bottle to create carbonation
 b. Larger and industrial operations instead carbonate the beer after the yeast has pretty much worked through all the sugars in the brew.
- (42) a. With your freezer example you won't see anything because the light turns off when you close the door... and the door is in the way.
 b. The light does NOT turn off.
- (43) a. [CMV: Trump's presidency is similar to putin's presidency during the mid 2000s] and this means eventually USA would become a corporate run oligarchy.
 b. Russia is not an oligarchy and definitely not a corporate-run oligarchy,

4.3 Answer (ANS)

Definition: a statement responding to a linked question.

Examples:

- (44) a. is it doing the exact same thing in Beer, or is the process any different?
 b. Exactly the same, yes.
- (45) a. Can a photon be detected that is not directly hitting a sensor of any kind?
 b. By definition, if it's detected then the device doing so is a detector,
- (46) Is anyone else feeling triggered by all the terrible things that are happening to women in Trump Abortion Ban states?
 (47) At this point, I feel more anxious and sick to my stomach than anything else.

4.4 Request for clarification (REQ)

Definition: a question or statement seeking further specification of the linked segment.

Examples:

- (48) a. The mechanism is the same as you described either way.
 b. is it doing the exact same thing in Beer, or is the process any different?

4.5 Continuation-IN (CONT_IN)

Definition: Segments linked to a previous segment within the same comment are systematically labeled as CONT_IN. For this reason, we do not provide examples.

4.6 Continuation-EX (CONT_EX)

Definition: a segment linked to a segment in another comment without signaling explicit agreement or disagreement.

Examples:

- (49) a. Buy an extender from a big box store to move everything out to where you need it.
 b. Thank you!
- (50) a. Essentially, every system of n stars is only 1 multi system but n stars with companions.
 b. Ok, I see what you mean,
- (51) a. But they could easily put text under the photo
 b. Studies are coming out that show that post Covid, reading is far less common among students.

4.7 Off-topic (OFF)

Definition: a segment that diverges from the topic of the linked segment. This means that there are not prior segment in the comment that is being replying to which coherently connect the replying segment.

Examples:

- (52) a. Then you're making beer.
 b. Vodka is fun.
- (53) a. The whole thing about letting them dry is only if you'll be using an oil-based coating.
 b. Not sure why this was downvoted before I upvoted it lol.

References

Harry Bunt, Volha Petukhova, Emer Gilmartin, Catherine Pelachaud, Alex Fang, Simon Keizer, and Laurent Prévot. The ISO Standard for Dialogue Act Annotation, Second Edition. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 549–558, Marseille, France, May 2020. European Language Resources Association. ISBN 979-10-95546-34-4. URL <https://aclanthology.org/2020.lrec-1.69>.