

## Annotation Guidelines for Formality of Reddit Titles

To annotate Reddit post titles from r/berkeley (the subreddit “berkeley”) based on the formality of post titles, we decided to follow a numerical rating system going from 1 to 7. The key factors we are assessing are as follows, in no particular order:

- Format:
  - Is the author writing the title as a full sentence?
  - If it’s a full sentence, is it an imperative sentence with an implied subject? (e.g. “Go to the library!”)
  - Or is it a partial sentence? (e.g. “Fixing the problem with Berkeley’s EECS funding,” “EECS16B Midterm”)
  - Or does it have a more defined headline like structure? (e.g. “The EECS Funding Crisis,” “The EECS Funding Crisis: Finding New Sources of Income,” “Why I think the EECS Funding Crisis is Good,” “Berkeley Professor says: “It’s time for more EECS money.””)
- Capitalization:
  - How does the author use capitalization?
  - Do they follow capitalization rules of standard English? (i.e. only capitalizing the beginning of sentences or, if a title, only capitalizing the first letter of relevant words)
  - Or do they use capitalization in more slang/modern ways? (e.g. all capitalization, spongebob mocking style  
<https://knowyourmeme.com/memes/mocking-spongebob>, etc.)
- Punctuation:
  - How does the author use punctuation?
  - Do they follow punctuation rules of standard English (i.e. only common and correct use cases of punctuation such as ‘.,!?’ etc.)?
    - Typically, use of exclamation marks, ellipses, question marks, etc. would be considered informal if unnecessary.
  - Or do they use punctuation in more slang/modern ways? (e.g. multiple exclamation marks or question marks, excessive use of ellipses, etc.)
- Grammar:
  - Does the author use correct standard English grammar or more slang/modern English? For example:
    - “My friends and I” vs. “Me and my friends” when subjects
    - Using “their” instead of “there” or vice versa
    - Use of double negation (“I don’t have none” vs. “I have none”)
    - Use of repeated words or other similar structures (e.g. “good good” vs “good”)

- Use of complex sentence structure vs. simple sentences (e.g. “As I went to the store, I discovered the sun was fading into the darkness” vs. “I went to the store”)
- Word choice:
  - Does the author use words in an official English dictionary, or do they use “unofficial words” (defined as emojis, profanity, slang, acronyms, shorthands, etc.)? If they use an acronym/shorthand do they include the full wording first? Do they use a combination?
    - CS (computer science) is a good example of it going both ways in that if we are talking about CS majors graduation, the University would say “Computer Science Majors” (more formal). But if we are talking about a specific class, say CS161, no one would say “Computer Science 161” (not necessarily less formal).
  - Even if the words are in an official English dictionary, are they appropriate for the context?
  - Do they include formal titles or honorifics? (e.g. Dr., Ph.D, etc., these would not be considered informal shorthands))
- Spelling
  - Does the author use only correct spelling?
  - Or do they intentionally misspell certain words? (this is related to word choice, e.g. “stonks”)
  - Or maybe unintentionally? (e.g. “berkley”)
- Topic:
  - Is the topic considered appropriate/serious, or, if inappropriate/not serious, is it discussed in a respectful way?
  - Our measure of “appropriate” is to what extent it could be discussed in a setting where behavior is expected to follow controlled standards, such as school or work.
  - Our measure of “serious” is to what extent would most consider to be important and not a joke.

### Rating system:

- 1: Extremely informal.
  - Partial sentence format
  - Constant use of incorrect capitalization, punctuation, grammar, misspelling, and/or unofficial words.
  - Topics may be inappropriate/not serious or discussed without respect.
  - Examples:
    - “fr tho”
    - “🙄🙄🙄”
    - “omg”
      - All are examples of “unofficial words” with no real sentence structure.
- 2: Very informal.
  - Partial sentence format
  - Repeated use of incorrect capitalization, punctuation, grammar, misspelling, and/or “unofficial words”.
  - Topics may be inappropriate/not serious or discussed without respect.
  - Examples:
    - “when ppl ask me which UC i go to”
      - The use of “ppl” and the capitalization of “UC” but not “i” leads to a very informal word choice and capitalization. UC is also an abbreviation.
    - “GO FREAKING BEARS”
      - All capitalization is informal, and then the use of the word “freaking” makes this one very informal.
    - “whoooooooo are you?”
      - Excessive use of extra letters in this intentional misspelling leads this to be very informal. Since it is also a proper sentence structure, it should also be capitalized.
- 3: Moderately informal.
  - Partial sentence/sentence format
  - Some use of incorrect capitalization, punctuation, grammar, misspelling, and/or “unofficial words”.
  - Topics may be inappropriate/not serious or discussed without respect.
  - Examples:
    - “Updoots for Cheeto the Physics Cat!”
      - “Updoots” is not a real word, and the use of exclamation marks with improper capitalization makes this informal.
    - “CAL BEARS BEAT STANFORD TO WIN THE BIG GAME 24-20”

- All capitalization but an otherwise proper sentence elads this to be informal.
  - “best feeling ever”
    - A partial sentence with no real structure (not a caption, headline or such) and also no proper capitalization or punctuation.
- 4: Neither formal nor informal.
  - Partial sentence/sentence format
  - Occasional use of incorrect capitalization, punctuation, grammar, misspelling, and/or unofficial words.
  - Topic is not extremely inappropriate/serious nor is discussed extremely disrespectfully.
  - Examples:
    - “When you realize you didn’t need winter clothes because you’re in California”, “The current state of this subreddit”
      - Partial sentences with correct capitalization, grammar, spelling, and no unofficial words.
    - “PSA to all students moving in Fall!”
      - Use of acronym and exclamation mark make this somewhat informal, but the structure otherwise is proper.
    - Many titles that fall in this category find themselves here because they may not fit in any other category.
- 5: Moderately formal.
  - Partial sentence format, sentence format or very defined format (see above for examples)
  - Majority of text uses correct capitalization, punctuation, and grammar, and/or spelling. Vast majority of words used are not “unofficial words.”
  - Topic is appropriate/serious and is discussed respectfully.
  - Examples:
    - “TIL: The median salary for a Berkeley lecturer is a whopping \$20K, and 66% of lecturers don't get health insurance.”
      - Topic is somewhat serious, there are some statistics mentioned. However, there is the use of the acronym “TIL”, abbreviation “K”, and a less formal word choice of “whopping”
    - “Missing this view from campus!”
      - Proper sentence structure with an omitted subject and exclamation mark makes this only moderately formal.
- 6: Very formal.
  - Sentence format or very defined format (see above for examples)
  - Vast majority of text uses correct capitalization, punctuation, grammar, and/or spelling. All/almost all words not “unofficial words.”

- Topic is appropriate/serious and is discussed respectfully.
- Ultimately, a 6 is a would-be 7 if not for 2 or 3 minor problems.
- Examples:
  - “Things are really hard for current bears. Who needs a Costco trip.”
    - The punctuation on the second sentence should be a question mark and the choice of “Things” creates a less formal sentence.
  - “Berkeley students, amid the 1918 flu pandemic, getting haircuts outdoors to avoid crowded spaces”
    - In this case, the post is most likely captioning a photo. Though it is a partial sentence, it still maintains correct spelling, punctuation, grammar, has no unofficial words, and is an appropriate topic that is appropriately discussed.
  - “MLK speaks in front of Sproul Hall (1967). Friendly reminder that Black Lives Matter, bears.”
    - MLK would ideally be first stated as “Martin Luther King Jr. (MLK)”, and the second part is not a full sentence when the first part was. But other than that, it has no other issues.
- 7: Extremely formal.
  - Sentence or very defined format
  - Fully correct capitalization, punctuation, grammar, and/or spelling. All words used are not “unofficial words.”
  - Topic is appropriate/serious and is discussed respectfully.
  - One potential error is allowed, but more than that would result in a 6.
  - Examples:
    - “Rest in peace, Courtney.”
    - “With two new apartment buildings opening this fall, UCLA will become the first and only University of California campus to guarantee housing for four years to first-year students and two years for transfer students.”
    - “Ms. Jules Means, Sociology '20, graduates at 67 with plans for an MSW. "It's never too late to achieve your aspirations in life.'”
    - “University of California Will End Use of SAT and ACT in Admissions”, “Dr. Fauci: COVID-19 vaccines should allow California colleges to teach in-person next fall”, “Kennedy at Berkeley's University House (1962)”, “War in Ukraine”
      - Use of headline structures with proper capitalization

Note:

We think the judgment factor in these decisions comes from the fact that not all categories are weighted equally in every case. For instance, unique capitalization can outweigh proper

grammar, spelling, etc if the title is like “dId I sTuTtEr?,” but capitalization is not necessarily always the most important category. Or perhaps, it has proper sentence structure, but the author uses somewhat informal words (e.g. “Least sexy college mascot in America? I beg to differ.”). But given a certain context, sometimes these words are not necessarily informal (e.g. “Cal Fiji mocking victims of sexual assault as they protest in front of their frat.”) Another good example is “Thank you, Berkeley, for letting me study at the school of my dreams :)”. It is worth noting that the author used double commas around “Berkeley,” which is notably proper grammar, but then used a smiley face emoji. This is all to show that there requires human judgment in these decisions.