

Annotation Guidelines

Our task attempts to solve the question: “Are ted talks funny?” We analyzed a dataset found on Kaggle that contained thousands of transcripts of previous ted talks and chose a random sample of 500 transcripts to analyze. We then broke down these transcripts into segments of the first 100 words to determine comedic relief. The first 100 words gave an overview of the talk and indicated the overall tone. The transcripts also include moments that indicate when the audience laughs at the talk (i.e. “(laughter)” or “(applause)” which helped us more accurately assess humor.

To classify our data, we included three categories:

- 1) Transcripts with a score of 1 are “not funny”.
- 2) Transcripts with a score of 2 are “neutral”.
- 3) Transcripts with a score of 3 are “funny”.

The score of 1 indicates the text is not funny and references serious and academic topics with little potential for humor.

The score of 2 indicates transcripts that utilize comedic relief but the laughter does not occur frequently, or transcripts where the topic leans more light-hearted and has potential for occasional humor.

The score of 3 represents the most humorous Ted Talks with frequent laughter and a light-hearted topic.

Some general guidelines we used to determine categories include: how much laughter from the audience, the number of jokes, the quality of jokes, and the tone of the text. We will now dive deeper into each category and take a look at some examples to learn how to determine borderline cases:

Score of 1: Not Funny

Transcripts in this category include serious topics related to illness, violence, poverty, death, feminism, immigration, etc. Any topics that fall under serious categories have no potential for humor so we mark them as “not funny”.

Additionally, texts with academic topics such as linguistics, science, historical events, etc are categorized as not funny unless they include clear jokes or indicate audience laughter. Texts that maintain a serious tone and use academic language fall under this category. The use of third person language most strongly indicates this category. However, many texts in first person also fall under this category.

Texts in this category have 0 jokes.

Example: “If there's any power in design, that's the power of synthesis. The more complex the problem, the more the need for simplicity. So allow me to share three cases where we tried to apply design's power of synthesis. Let's start with the global challenge of urbanization. It's a fact that people are moving towards cities. and even if counterintuitive, it's good news. Evidence shows that people are better off in cities. But there's a problem that I would call the "3S" menace: The scale, speed, and scarcity of means with which we will have to respond to this phenomenon has no”

Score of 2: Neutral

Transcripts in this category may cover a variety of topics including: personal anecdotes, academic topics, philosophical questions, etc.

While some topics may be similar to those in the “not funny” category, texts in this category maintain a conversational tone whereas those with a score of 1 use formal language.

Texts that have “(laughter)” denoted but have an overall academic tone fall into this category.

Additionally, the use of first person can be an indicator the text is neutral. While the topics may be more academic/science-related, first person shows jokes are more likely to be told. Texts in this category have 0-1 jokes.

Example: “For the next 16 minutes, I'm going to take you on a journey that is probably the biggest dream of humanity: to understand the code of life. So for me, everything started many, many years ago when I met the first 3D printer. The concept was fascinating. A 3D printer needs three elements: a bit of information, some raw material, some energy, and it can produce any object that was not there before. I was doing physics, I was coming back home and I realized that I actually always knew a 3D printer. And everyone does. It was my mom. (Laughter) My mom takes”

Score of 3: Funny

Transcripts in this category do not include any serious topics mentioned in the “not funny” category and typically are personal anecdotes, light-hearted topics such as sex, music, and family/children. Topics often are related to science, but the inclusion of multiple jokes and denoted “(laughter)” show that the tone of the talk itself is light-hearted.

These texts are always told in first person.

Texts in this category have at least 1 joke.

Example: “So, people are more afraid of insects than they are of dying. (Laughter) At least, according to a 1973 "Book of Lists" survey which preceded all those online best, worst, funniest lists that you see today. Only heights and public speaking exceeded the six-legged

as sources of fear. And I suspect if you had put spiders in there, the combinations of insects and spiders would have just topped the chart. Now, I am not one of those people. I really love insects. I think they're interesting and beautiful, and sometimes even cute. (Laughter) And I'm not alone. For centuries, some"

Incidental occurrences:

- Transcripts with musical lyrics (denoted by the 🎵 symbol in the text) are assigned a score of 1.
- Blank transcripts are assigned a score of 1.

Borderline Case Example: Category 1 vs 2

'If we look around us, much of what surrounds us started life as various rocks and sludge buried in the ground in various places in the world. But, of course, they don't look like rocks and sludge now. They look like TV cameras, monitors, annoying radio mics. And so this magical transformation is what I was trying to get at with my project, which became known as the Toaster Project. And it was also inspired by this quote from Douglas Adams, and the situation is from "The Hitchhiker's Guide to the Galaxy." And the situation it describes is the hero...'

We marked this text as "neutral". While the topic is academic and related to science, the use of first person and casual tone indicates that there could be potential for jokes later in the talk. The use of the adjectives "magical" and "annoying" indicate the casual tone.

Borderline Case Example: Category 2 vs 3

'Puzzles and magic. I work in what most people think are two distinct fields, but I believe they are the same. I am both a magician and a New York Times crossword puzzle constructor, which basically means I've taken the world's two nerdiest hobbies and combined them into one career. And I believe that magic and puzzles are the same because they both key into one of the most important human drives: the urge to solve. Human beings are wired to solve, to make order out of chaos. It's certainly true for me. I've been solving my whole life. High school...'

We marked this text as "funny". The topic is related to puzzles/general philosophy, showing it could fall into category 2 or 3. However, the joke "which basically means I've taken the world's two nerdiest hobbies and combined them into one career" and the casual language shows the tone of the text will be light-hearted and show potential for future jokes throughout the transcript.