Humor Classification Annotation Guidelines

Background and Context:

You will be annotating a collection of tweets entered into the #HashtagWars competition from the comedy central show @midnight. The competition was organized as follows. Each episode of the show proposed a topic in the form of a Twitter hashtag, viewers were asked to enter a tweet into the competition using the hashtag and the next episode the show ranks the provided tweets based on humor. The tweets in the dataset are each annotated with a score, outside of the score, the relative humor rankings in the dataset are undifferentiated. The score for the winning tweet in a hashtag category is 2, every other tweet in the top ten has a score of 1, and the remaining tweets in the hashtag category receive a score of 0.

Task:

Your task is to label each tweet in each hashtag category with the categories it satisfies from the provided humor taxonomy. The categories in the humor taxonomy are:
1. Absurdism
2. Insult
3. Irony
4. Observational
5. Wordplay
6. Vulgarity
7. Other
8. Not humor

This taxonomy is a revision of the twelve category taxonomy given in Hay 1995[1], absurdism being our only addition not built out of pre-existing categories in the paper.

---

[1] https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.3.5937&rep=rep1&type=pdf

Identifying Type of Humor:

Absurdism:

Instances of this category of humor tend to operate by deliberately disrupting or subverting causal reasoning or expectations. This differs from other discrepancy based humor like irony since irony is indicated by discrepancies in present multiple meanings of an utterance. Absurd humor can also contain humor based on awkward or uncomfortable states of affairs. For example:

695345270792413184   Chris Hardwick anime body pillow #BadKickstarterIn5Words @midnight   0

While this is humorous, it does not seem to involve irony, insult, observational, or wordplay. While anime body pillows are perhaps taboo in that they are generally not an admissible topic of conversation, this does not seem to fit the criteria of vulgar humor in that they generally aren't considered obscene, nor does this tweet's humor seem to come from shock value. The fact of the content's strangeness or awkwardness seems to be the source of the humor, and in a way that doesn't correspond to insult or irony, since there is no self deprecation, insult to Chris Hardwick, or multiple notably differing meanings. The strangeness and inappropriateness of the situation described is the source of humor and thus this would be tagged as absurdism.

Main Requirements - Absurdism:
- **Situation or idea which disrupts expectations or reasoning**
- **Potentially awkward, uncomfortable, baffling, or "cringe"**

Insult:

Hay outlines insults in humor as follows: "An insult is a remark that puts someone down or ascribes a negative characteristic to them There are two distinct types of humour which come under this heading One is jocular abuse in which the speaker jokingly insults a member of the audience The other is when someone absent is insulted The insult here is likely to be genuine and the humour stems from the unexpectedness of the statement which in most circumstances would be unacceptable." (70) It is possible that the show @midnight itself might be the target of an insult, but most likely insults will take the second form described, for example:

674809415976009729   I won't vote for Trump for the sake of doing a bit. #RealisticResolutions #PointsMe @midnight   1

Self deprecating humor, where the speaker or author ascribes a negative characteristic to themself will be counted as insult, since the content of a tweet can be an insult targeted towards the author.   According to Hay this has the potential to be used as a "defense strategy" (78), since if the author points out a flaw before anyone else, any further insult regarding the flaw will have less force. An example of a self deprecating insult tweet would be:

674818891772780544   Gonna be less passive aggressive as if anyone would notice #RealisticResolutions @midnight #PointsMe   1

NB: The above tweet qualifies as ironic humor as well since the author intentionally displays the negative characteristic they ascribe to themselves.

Main Requirements - Insult:
- **There is a word/phrase establishing identity of the target of the insult.**
- **There is a negative comment/association.**

Irony:

There are numerous ways to define irony, one of the easier and more intuitive ways is given by Hay: "If the speaker does not mean their words to be taken literally and in saying them is implying the opposite or something with a markedly different meaning then it is classed as irony." (71) Irony depends both on the speaker and the context, however since the tweets you are annotating are read without taking the author into consideration mostly you will be evaluating this with respect to the content of the tweet along with its relationship to the hashtag. For example:

674809212262830081   Will only eat carbs, proteins and fats #RealisticResolutions
@midnight #PointsMe   1

The above tweet is classified as irony since while the content of the tweet on the surface level indicates a desire to restrict their diet for health, all the contents of food fall into either the category of either carbohydrates, fats, or proteins, so despite the supposed intent towards restriction, what is described does not constitute a restriction.

Main Requirements - Irony:
- **Consider what the speaker truly believes/means.**
- **Consider what is actually said.**
- **Check if there is a notable discrepancy between them.**

Observational:

Observational humor involves either a remark or insight about something humorous in the relevant context, or environment, or making a humorous statement out of something in the relevant context or environment. According to Hay observational humor"consists of quips or comments about the environment the events occurring at the time or about the previous person's words. In this sense it is observational the speaker is making an observation about something funny or making a witty observation." (74) Since the tweets in the dataset do not refer to other tweets, you will not need to consider the possible instances of observational humor which rely on other tweets in the discourse/relevant hashtag category. For examples:

673517587213344769   Here comes Santa Claus, here comes Santa Claus.. Oh shit he driving on the side walk @midnight #DrunkCarols #PointsMe   0
717558924073951233   Brittle Women #ElderlyMovies @midnight   2
694771099037356033   #BeforeFacebookI thought my kid was special. @midnight 1

The first tweet is an instance of observational humor since it involves a known or possible example of drunken behavior which relates the content of the tweet to the hashtag. The second is a case of observational humor since it references the fact that often as people age their bones tend to become more brittle, given this tweet is also playing on the phonetic similarity between little and brittle, this also fits the category of wordplay. For the third tweet, the humor is due to the fact that social media use often changes people's perception of themselves and others.

Main Requirements - Observation:
- **There is a word/phrase making a reference to outside knowledge.**
- **In particular, there is an event, person/character, work (movie, song etc), quote, stereotype, etc. that is being referenced.**

Wordplay:

According to Hay "Wordplay is any humorous statement in which the humour derives from the meanings, sounds or ambiguities of words. The most typical instance would be a pun. This may involve a speaker deliberately punning or a hearer identifying an ambiguity in the speaker's speech and exploiting it for humorous effect." (79).   For example:

723186463936913408   Jordan Sparks a fatty #420Celebs @midnight   0
721025637968846848   @midnight Sgt Pepper's Lonely Darts Club
#AddSportsRuinASong   0

The first of these tweets' wordplay is based on the last name 'Sparks' and the verb 'sparks' being homonyms, the verb here having a use characteristic to lighting cannabis cigarettes and the name belonging to a celebrity, both of these relate the content to the hashtag. In the second, 'Sgt Pepper's Lonely Hearts Club Band' is an album, and 'Darts' is a sport, phonetically similar to the word 'Hearts', again, these two things relate the content to the hashtag.

Main Requirements - Wordplay:
- **There are word(s) are exploited for humorous effect on the basis of ambiguity in sound, meaning, or orthographical form.**
- **These can be different types, like pun, palindrome, spoonerisms, etc.**

Vulgarity:

Vulgar humor derives from "derives from the fact that the speakers are breaking some sort of taboo" (Hay 79). In practice this typically manifests as toilet humor, sexual humor, or unexpected use of obscenity relating to these types of humor, which acts as a source of shock or incongruity within the context. Examples:

656673795122556928   #HitOnYourMom @midnight Can i stay the night in my old womb?   1
712472017128132608   @midnight goddamn racist mysoginist homophobic fuckfaces! #TwitterIn5Words   0 (vulgarity expressing extreme emotion)

For all our sakes we won't explicate the multiple implications contained in the first tweet, suffice it to say most of its humor predicates on the shock of flouting (multiple) taboos. This first tweet is also a case of a tweet fitting multiple categories, the other category being wordplay, since "womb" sounds similar to "room". The second tweet's humor comes from its shock in strong emotion from of its condemnation of twitter and the obscenity at the end ("goddamn" while still rarely if ever used in polite conversation, is perhaps less of an obscenity rather than an indicator of harsh emphasis which is only appropriate in certain contexts.)

Main Requirements - Vulgarity:
- **There is a word or phrase that is vulgar by definition (like curse words), or that is a source of shock in the current context.**

Other:

While Hay's taxonomy and our addition are grounded in empirical data there are still instances of humor which do not fall into the above categories or their beginnings in Hay. In Hay's findings only %5 of the data fell into the category 'other', but it will still likely be encountered here. If you encounter a tweet which makes a successful attempt at humor but does not fit into the other categories label it as 'other' and there will be a text field for you to list the components that you find contribute to the tweet's humor.

Not Humor:

These are for tweets that don't make an attempt at humor, usually ones that unsuccessfully flout the topic given by the hashtag. If a tweet does not fit the typical humor categories, and is not clearly funny enough to make it into 'other', you should label the tweet as 'not humor'. For example:

676855647271591936   #GOPSongs I wonder why @midnight DIDN'T do #MuslimSongs? (Oh, yeah, that's right!) #SouthPark #hypocrisy https://t.co/38MrTHZH8C 0

This tweet does not make an attempt to participate in making humorous content which relates to the hashtag in a typical way. It tries to make a humorous, critical comment on the contest itself, perhaps using irony and meta-humor, but for now that is beyond the scope of taxonomies of humor we considered.

Annotation:

To start annotating tweets, you'll need [MAE](). Included alongside this document is the DTD for annotating for this task with MAE. Within MAE, load the DTD as a new task definition, open the first .xml file in the folder assigned to you, and follow these steps for each tweet in the file:

1. In the top half of the MAE interface, select the text span corresponding to the ID number of the tweet, then right click the span and select "Create Extent Tag with selected text" > "tweet". This tag will serve as the base for all category annotations for that tweet.
2. Read the tweet and decide on a category of humor that is present in the tweet.
3. In the "Tags" menu, click "Create NC Extent Tag with no text associated" and select the category you found in step 2.
4. The tab for the category you selected will open in the bottom half of the MAE interface. Find the tag you just created and add your notes for how the tweet meets the requirements for the category in the "notes" field.
5. Switch to the "All Extents" tab in the bottom half of the MAE interface and find the tweet and the NC Extent Tag you just created
6. Using the Control key (Windows, Linux generally) or the Command key (Mac) to select multiple extents in the list, select the two tags.
7. Right click one of the selected tags and click "Create Link tag with selected arguments" > "Humor"
8. Repeat steps 2-7 for any additional categories of humor that you decide are present in the tweet.

See the video included with the annotation package for a demonstration of the above steps. *There is no minimum or maximum number of humor categories for a single tweet. In fact, multiple types of humor are often interwoven. For example, a common pattern is Observational and Wordplay humor when making a pun on the name of a song, movie, etc., since it requires knowledge of the work and recognition of the pun. The annotator should mark tweets for all perceivable forms of humor.

Works Cited: