

Annotation Instructions

In the following, you are asked to annotate a set of expressions. We ask you to decide for every individual expression whether you think it can be perceived as **offensive** or **not**. In case you do, we ask you to provide a sentence in which the given expression occurs that illustrates that offensiveness.

By offensive utterances we understand (unfounded) accusations, threats, general insults, profanity, or utterances generally considered rude.

Important:

You are to make a decision on the basis of whether you think that *someone* may perceive such an utterance (that you conceive with the given expression) as offensive/insulting. We are **not asking whether you personally would feel offended or insulted**. You may be toughened up or thin-skinned. Try to give a judgment that you think would be representative of the general public. However, **if you are not entirely sure whether an expression is offensive, you'd better regard the word as offensive**. Regard yourself as a *watchdog* whose job is to identify authors of offensive utterances and thus prevent future attacks from these wrongdoers.

We will randomly present you either a verb, a noun or an adjective. Below we list for each part of speech likely contexts for offensive expressions. We would also like you to use these lists of phrases for the expressions given to you. If you think that the given expression is offensive but it does not fit any of those phrases, then you may use an existing phrase and modify it. However, **please try not to alter these phrases significantly**.

Reference contexts for **nouns** (note that in each utterance we mark the noun under consideration in **bold**):

- You are a **coward**.
- You are responsible for this **massacre**.
- We don't want your **selfishness** here.
- I really can't stand your **hypocrisy**.
- I really can't stand your **meanness**.
- I really can't stand your **stubbornness**.

Reference contexts for **adjectives** (note that in each utterance we mark the adjective under consideration in **bold**):

- You are so **treacherous**.
- You are so **lazy**.
- Don't be so **stupid**.
- Don't be so **foolish**.
- You are such a **useless** being.
- You are such an **obtuse** person.

Reference contexts for **verbs** (note that in each utterance we mark the verb under consideration in **bold**):

- You are **gossiping** all the time.
- I simply **laugh** at you.
- You **sicken** me.
- Stop **boasting**.
- I will **kill** you.
- You just **screwed** everything.
- You **besmirched** my entire family.

Please bear in mind the following **advice**:

- All expressions to be annotated bear a **negative** connotation. However, **by far this does not mean that every expression is also offensive**. Even if the given expression could produce a well-formed English sentence if inserted in one of the above reference sentences, it does not inevitably mean that this expression conveys offensiveness. For instance, *disappointed* is unlikely to convey offensiveness as part of the sentence *Don't be so disappointed*. Similarly, *critic* does not convey offensiveness despite the sentence *You are a critic*.
- Please remember that the **major goal of this annotation is to isolate the offensive expressions** among the list of negative expressions given to you. The sentences you provide are just some form of sanity check. **Only provide a sentence when you think that the present negative expression is offensive**. If you provide a sentence for every given expression in this survey, then you obviously misunderstood the task.
- **The offensive nature of a sentence must clearly originate from the given expression**. Therefore, we ask you to stick to the reference phrases from above as closely as possible, as they should ensure this).
- The example sentences you produce must be **well-formed English sentences**. You should be convinced that other native speakers of English would actually produce such sentences.
- We assume that one prerequisite of conveying offensiveness is that the given expression needs to relate to some person – typically the 2nd person *you* (all of the above reference sentences do). Therefore, if you cannot find any context for a given expression where there is such a relation, then the expression is less likely to convey offensiveness.
- In case, you feel you must alter the phrases, your changes should only be minimal. Also check that the offensive content does not originate from an expression in the sentence other than the target expression. For example, if you were to classify the word *disappoint* then the sentence *You disappoint me, you bitch*. would be inappropriate since in that sentence the insult is not conveyed by the word *disappoint* but the word *bitch*.
- Sometimes, the entire event described by a sentence is offensive but the words in isolation are not. For example, the sentence *You petrify at the sound of a mouse*. can be perceived as an offense. However, it is not the negative word *petrify* which causes that perception but the image of a coward that is conjured up by the entire sentence. We are not interested in words, such as *petrify*.

In the following, we give some examples which we consider not offensive and explain them (these examples also give an idea in how far we allow the modification of the above reference phrases):

- I can't stand your **criticism**. This utterance may be considered unusual. Moreover, criticism is no offensive action. Things would be different, though, if instead of

criticism, nouns such as *nagging*, *harping* or *squabbling* were used. (For instance, if someone says that you are nagging, (s)he challenges your ability to provide an adequate form of criticism -- if your actions are described as criticism there is no such pejorative connotation that would make you feel hurt.)

- You are a **victim** of domestic violence. Such a statement suggests that the speaker feels pity for the person (s)he refers to as a victim.
- We don't need your **warning**. A warning is no offensive action. (On the contrary, often you warn someone in order to protect her/him from danger.)
- We don't want your **regret**. (Again, a slightly unusual utterance.) Regret is a frame of mind which has no offensive connotation.
- You are responsible for this **mistake**. Too mild in order to be perceived as offensive.
- Don't be so **irritated**. This utterance could be made to calm someone down. An offensive intention is less likely.
- You are so **unhappy**. This utterance may also be considered unusual. In any case, *unhappy* describes the frame of mind of the addressee (i.e. *you*). There is nothing inherently offensive about being sad, upset or disappointed etc.
- Don't be so **reluctant**. The speaker may want to criticize the person addressed. However, an offense is unlikely to be intended by this remark. Besides, reluctance is not a negative human property per se.
- You are a **restless** soul. Being restless may have both a positive and negative connotation. Even if you think that the negative connotation prevails, this property is still too mild to be considered offensive.
- You are so **inappropriate**. This utterance may be considered unusual. Moreover, *inappropriate* is a fairly polite word.
- I **protest** against you. Protest is not inherently offensive (this is similar to *criticism*).
- Stop **arguing**. This is an order which typically a parent gives to his/her child. (It may be inappropriate to give such an order to other persons, e.g. your boss -- but the order as such is not offensive.)
- You **startle** me. One can be startled by both positive and negative things. Even if you think that typically negative things are meant, then this utterance could still be considered as just some form of criticism.
- I will **forget** you. There may be contexts where such an utterance is acceptable but most of them would be fairly contrived.

Our recommendation for the annotation is that you store the above guidelines or print them out. You should **use the above examples as a reference** for your annotation.