Supplementary Notes to *Implicitly Abusive Comparisons – A New Dataset and Linguistic Analysis*

January 20, 2021

1 Introduction

This document provides more detailed information regarding certain aspects of our research for which there was not sufficient space in the main paper. We focus on the following aspects:

- background information regarding the design of the crowdsourcing tasks (§2)
- some words on the acceptability of individual comparisons (§3)
- annotation guidelines for manually created features (§4)
- configuration of supervised classifiers (§5)

2 Crowdsourcing Tasks

Figure 1 illustrates the data collection process that we devised in order to produce the final dataset of 1000 comparisons with the two classes being *abusive* and *non-abusive* comparisons. For the exact guidelines that were used, we refer the reader to the supplementary material which include the pdf files that we had given to the crowdworkers. These pdf files were provided by an external anonymous URL enabling the crowdworkers to browse them while working on the task. This helped crowdworkers clarify open questions that arose while working on a task.

The comparisons were created in two different tasks, one was dedicated to creating abusive comparisons, the other was dedicated to negative non-abusive comparisons. Figure 2 illustrates a typical question page of the former task while Figure 3 illustrates one of the latter task. Since we noticed that the crowdworkers had more difficulties with non-abusive comparisons that are still negative in polarity[1], we provided them with situations that we thought were typical contexts in which non-abusive comparisons would naturally arise. In Figure 3 one of those situations is stated (the crowdworkers are to express their

[1]The crowdworkers tend to create positive or neutral comparisons when asked to invent non-abusive comparisons.

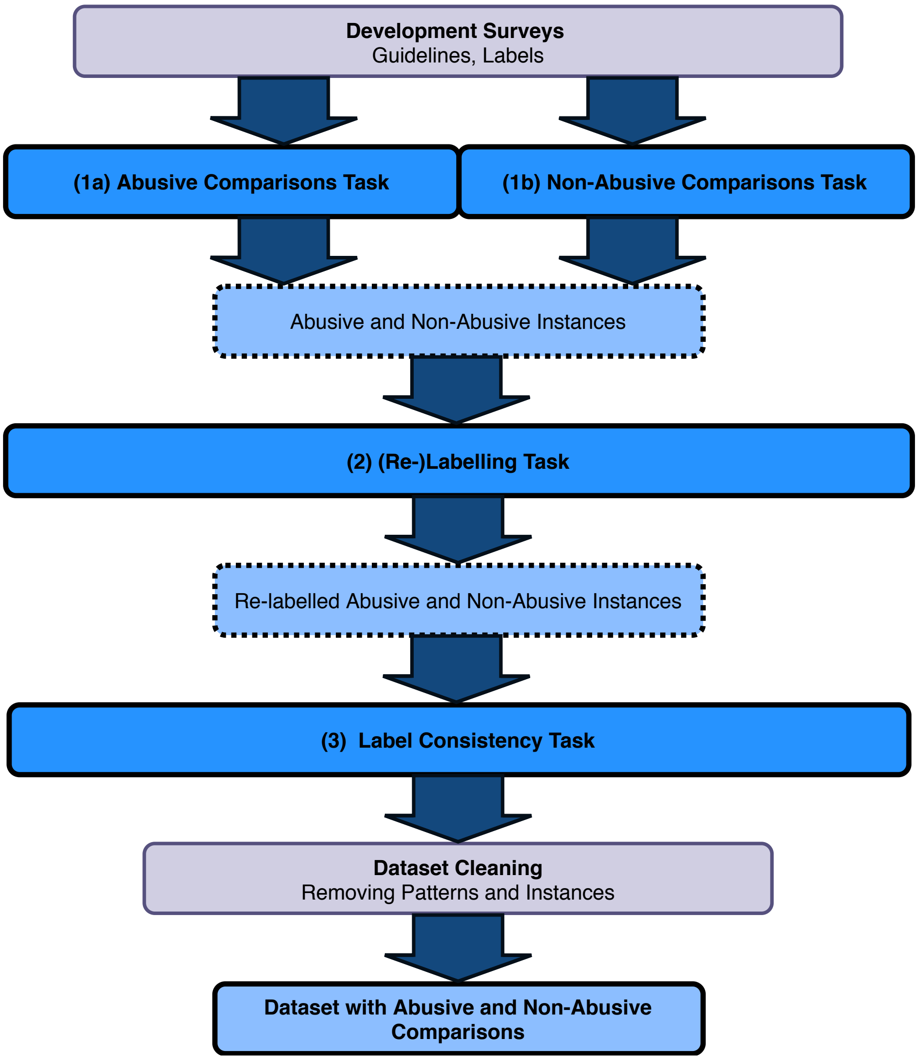

Figure 1: Overview of the data collection process.

concern at how ill they think their colleague is at work). In total, we used 45 unique situational frames. A complete list of the situational frames is included in Appendix A.

Since we wanted our comparisons to be context-independent (that is, our supervised classifiers should classify them without knowing about the situation that was provided to the crowdworkers), we added a further step in which additional crowdworkers had to validate the label of the comparisons that were created in the previous step in the absence of the given situation. In order not to bias the crowdworkers, the comparisons that a crowdworker had to assess did not include the original label of the previous task(s). Figure 4 illustrates a typical question page from this re-labelling task.

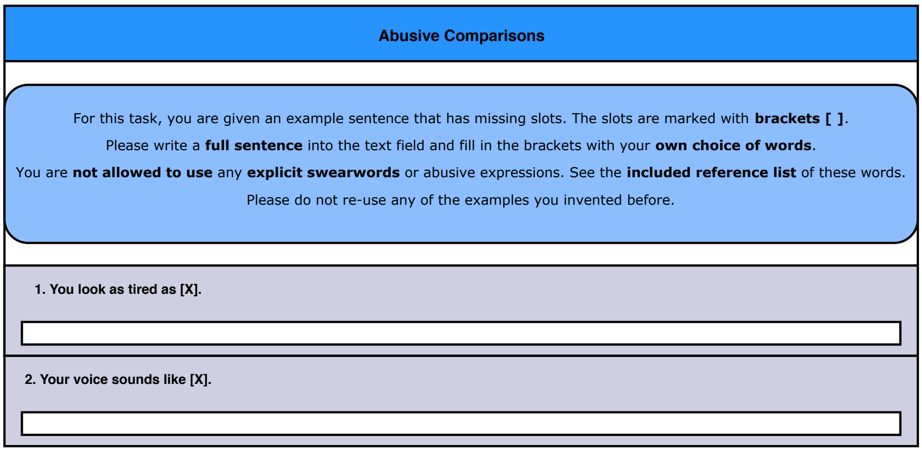

Figure 2: Example question from the task of creating abusive comparisons.

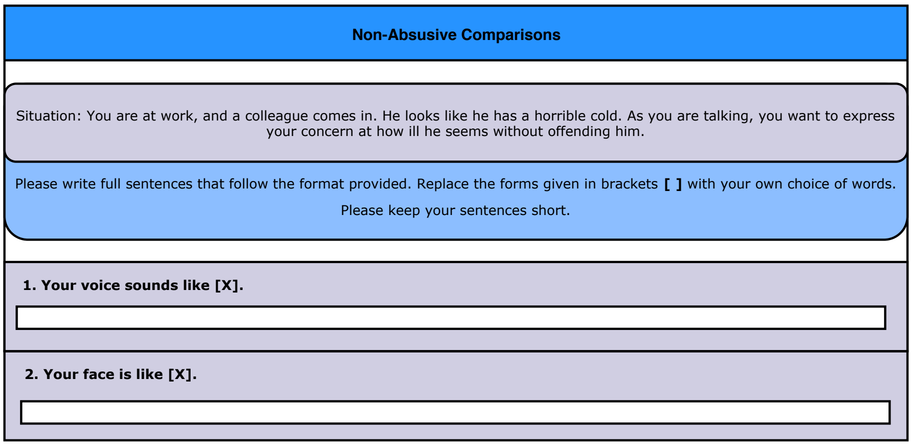

Figure 3: Example question from the task of creating non-abusive comparisons.

In the final step, instances with semantically similar content but different class labels were re-checked by additional crowdworkers. Theoretically, there may be two reasons for these different labels: either the comparisons really do have different labels, or the labels that had been previously assigned were inconsistent. In the latter case, the instances of a group should be assigned the same class label. (Inconsistencies may arise since in the re-labelling task, the comparisons to be re-labelled are distributed among dozens of crowdworkers, so that a single crowdworker may only re-label a subset of the total set of comparisons. Different crowdworkers may have different thresholds for categorizing an utterance as abusive.) Figure 5 illustrates a question type from this label consistency task.

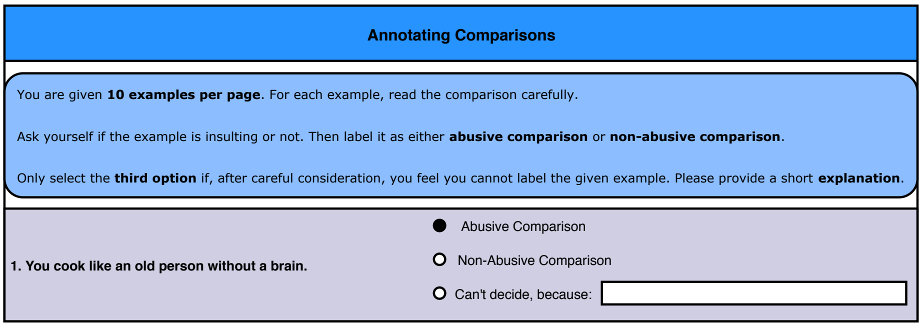

Figure 4: Example question from the re-labelling task.

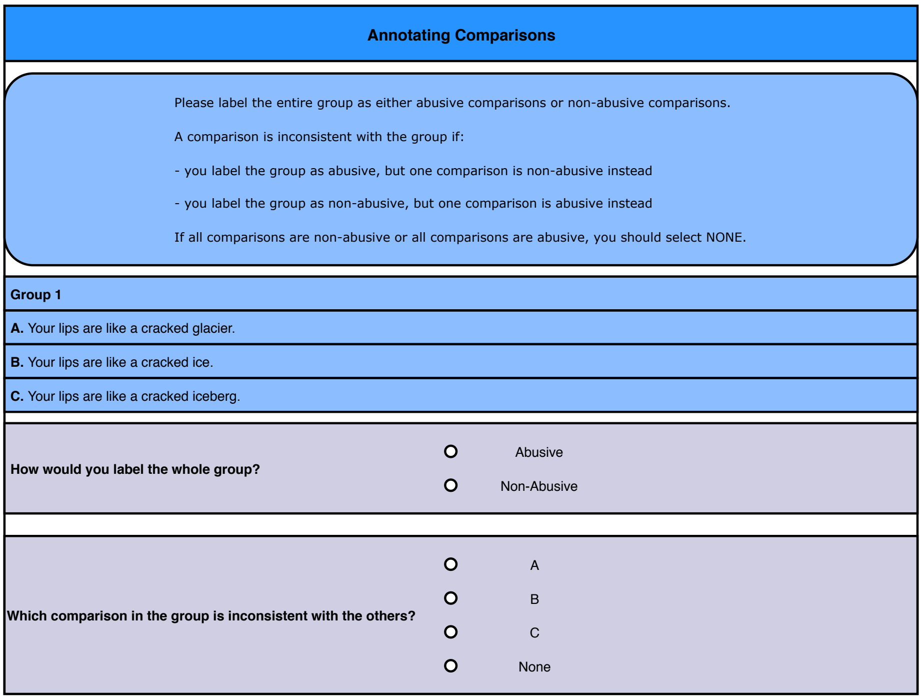

Figure 5: Example question from the label consistency task.

3 Some Words on the Acceptability of Individual Comparisons

Inventing abusive and non-abusive comparisons is generally a difficult task. Given that our crowdworkers are effectively laypersons, the quality of their invented comparisons varies. This raises the question what criteria were defined that decided whether a particular comparison was included in our final gold standard.

The re-labelling task was designed in such a way that it also allowed crowdworkers to label a particular comparison as *can't decide* (Figure 4). If the language of a comparison was too poor, its meaning was unclear or the language was too ambiguous so that the crowdworkers were unsure whether the utterance was meant to be abusive or not, then typically crowdworkers should make use of this label. Consequently, that particular comparison was excluded from the final gold standard.

Despite this option in the re-labelling task, the authors of this paper additionally inspected each comparison manually to guarantee that problematic comparisons described above were also excluded in case the crowdworkers were too lenient. This was done as part of our *dataset cleaning* step of our data collection process (Figure 1). Already prior to the re-labelling task, the authors made sanity checks on the crowdworkers' responses from the task of creating abusive and non-abusive comparisons. In the unlikely case that a crowdworker would always resort to the same type of vehicles (e.g. *You move like a storm*, *Your hair looks like a storm*, *You are as organised as a storm* etc.), such response would already be excluded from further processing at this stage. However, we only recall very few of such incidences. The set of requirements that crowdworkers had to fulfil in order to be eligible for one of our tasks was also fairly strict (basic academic education, such as a Bachelor's or Master's degree; an approval rate of 90% or higher; no dyslexia).[2]

Some of the comparisons invented by our crowdworkers described images that are not perfectly clear, such as (1) or (2). We kept such comparisons in our dataset in case it was still clear that these comparisons were meant to be abusive or not. (For example, (1) and (2) were kept since they were considered not to be meant abusive.)

(1) Your presence is like a dark stormy night.

(2) Your words are like fire.

The decision for keeping such comparisons was that we wanted to have a dataset representative of real language. By inspecting comparisons on social-media sites, we found that there, too, the images that are described are not always perfectly clear.

Of course, the overwhelming majority of comparisons in our final dataset are comparisons that did not display the above shortcomings.

In general, all comparisons of our dataset were meant to have a negative polarity. This is because, the task of polarity classification on comparisons has already been addressed in previous work [Qadir et al., 2015]. Comparisons produced by our crowdworkers that did not match that polarity, i.e. positive or neutral comparisons, were excluded from the final dataset. However, we were lenient when a comparison was ambiguous with regard to its polarity. For example, (3) and (4) could be interpreted as negative polar remarks but one could also imagine special contexts in which they are not meant that way.

(3) Your eyes are like military grade lasers.

[2] *Prolific Academic* has minimum payments ($6.5 per hour) much beyond other crowdsourcing platforms, such as *Amazon Mechanical Turk*. This typically results in high-quality responses.

(4) You dance like no-one is watching.

We kept all comparisons with ambiguous polarity in case we could imagine a plausible context in which those comparisons would be used in a negative way. In other words, our dataset will also contain a few comparisons that are only potentially negative. Since the automatic classification of polar comparisons is far from perfect [Qadir et al., 2015], we think it is reasonable to work with a dataset which contains the most realistic set of human-produced comparisons that systems predict to be negative. If such a classifier is tuned for high recall, comparisons, such as (3) and (4), are likely to be contained in the set of assumed negative instances. This procedure follows the argumentation of the *second classification step* in Wilson et al. [2005]. Our dataset may include a few comparisons too many. But only thus can we be fairly certain to account for those instances that would need to be classified under realistic circumstances.

Some of the comparisons in our dataset may sound less fluent although they are grammatically correct and the imagery used is easy to follow. Typically, the lack of fluency was caused by enforcing the usage of the given patterns, such as in (5) and (6).

(5) You are as hungry as a marathon finisher.

(6) Your conversations are like walking through treacle.

As we have shown in the main paper, it is vital to provide the patterns to the crowdworkers as, otherwise, unforeseen biases in the language of the comparison may be introduced (coincidentally). This usually results in overly optimistic performance scores in supervised classification. However, to acknowledge that the odd comparison may sound slightly construed, we produced a **second (revised) version of the dataset** (*it is also contained in the supplementary material*) in which some language deficiencies were manually corrected. For instance, (5) and (6) were replaced by (7) and (8). In this second version, we also corrected grammatical and orthographic mistakes.

(7) You look as hungry as a marathon finisher.

(8) Talking to you is like walking through treacle.

Our experiments as reported in the main paper, however, were all done on the original dataset. We did so since, with those language shortcomings such comparisons are not only more difficult to process but also more representative of real-life language, particularly on social-media sites.

As a test, we also ran our best supervised classifier (i.e. BERT with linguistic features) on this second (revised) version of our dataset. In terms of classification performance, there was no notable difference between those two versions.

4 Annotation Guidelines for the Manually Created Features

The guidelines for the more difficult manually created features, i.e. the distinction between figurative and literal language and the detection of dehumanization, were already described in detail in the main paper. There now follows some additional information on the other manually extracted features.

4.1 Contradiction

By contradictions, we understand constructions where the property of the comparison is opposite to the prototypical properties associated with the vehicle. (9) is a contradiction since a caveman is the epitome of outdatedness. Therefore, it is definitely not modern. (10) is a contradiction because a monkey, being an animal has no manners. So it has no refined manners for sure. Contradictions may not always rely on some adjective as the previous examples suggest. For instance, in (11) the contradiction is between the noun *progress* and the event *stepping backwards* which are two opposing concepts. Some comparisons may be conventionalized and represent idioms, such as *as clear as mud* (12). We do not make a distinction between these conventionalized and the other contradictions simply because contradictions are fairly rare in general on our dataset (3.2% of the comparisons).

(9) You are as modern as a caveman.

(10) Your manners are as refined as a monkey's.

(11) Your progress is like you're stepping backwards.

(12) Your words are as clear as mud.

Any contradiction that is the result of linguistic shortcomings is <u>not</u> labelled as a contradiction. With this feature, we only consider those contradictions which are deliberately conceived by the author of the comparison. It means that they are meant as a stylistic device. This is true of all of the above examples. Such instances are typically very easy to spot manually, particularly since many of the cases in our dataset are idioms (12), which explains the perfect interannotation agreement we measured between our two expert annotators on a sample of 200 comparisons (see Table 4 in the main paper).

4.2 Absurd Images

For this feature, only the vehicle was considered since the remaining components of the comparison do usually not have a figurative meaning. By *absurd* image we define any event that never comes true (13) or is at least highly unlikely (14). We also included actions that are, in principle, not impossible but if they are carried out, one would question the mental sanity of the person who performed that action (15). Apart from that we included idioms by this category if the idiom literally describes an absurd event (16).

(13) Your input is like a baby giving their opinion on computer code.

(14) Your approach is like a bus crashing into a oil tanker.

(15) You cook like you read the instructions backwards.

(16) Your manners are like a bull in a china shop.

Any comparison that is considered vague or unclear by the annotator was not considered as an absurd image. The absurdity should be intentional and not be the result of linguistic shortcoming on the part of the author of that comparison.

4.3 Evaluation vs. Emotional Frame of Mind

As already stated in the main paper, the distinction between emotional frames and evaluations is almost synonymous to sentiment views [Wiegand et al., 2016] which divides sentiment into *actor views* (≈emotional frames of mind) and *speaker views* (≈evaluations). A major difference between sentiment views and emotional frames/evaluations is that our comparisons are not restricted to (verbal) sentiment. For example, there are quite a few images that involve animals. Such creatures are incapable of expressing sentiment (verbally). Therefore, the concept of sentiment views does not fit strictly speaking. However, emotional frames are a property that may also apply to animals, such as in (17) and (18).

(17) You walk like a scared penguin walking towards the rough sea.

(18) Your reaction reminds me of a shocked cat.

Regarding annotation guidelines, we refer the reader not only to [Wiegand et al., 2016] but also Wiebe et al. [2005] who initially proposed this distinction while creating the MPQA corpus. In that work, the terms *direct subjectivity* (=actor view, emotional frame of mind) and *expressive subjectivity* (=speaker view, evaluation) are used. However, that work does not address the annotation of individual words as Wiegand et al. [2016] and we do. Therefore, we primarily cited Wiegand et al. [2016] in our paper.

The manual annotation of our comparisons with regard to these categories was necessary despite the existing lexicon from Wiegand et al. [2016]. One reason for doing a manual annotation was that this lexicon seems inaccurate when it comes to ambiguous words. Wiegand et al. [2016] assign one category for each word thus assuming one sense per lexical entry. However, there are common words, such as *sick*, which may convey an evaluation in one context (meaning *crazy* or *mad*) or a judgment on the frame of mind (feeling bad as a result of suffering from illness) in another context. Another shortcoming of the existing lexicon was that our set of comparisons also contains several idioms, such as (19) or (20). The lexicon by Wiegand et al. [2016] only comprises unigrams and therefore all idioms observed in our dataset are missing.

(19) You are acting like you are *carrying weights on your back.*

(20) You reacted like you *got caught with your trousers down.*

5 Configuration of Supervised Classifiers

Our focused dataset with 1000 instances is small compared to other general datasets currently available for abusive language detection. As we did not have a dedicated development set, we decided to employ only those classifiers that either require no or only little hyperparameter tuning (such as fastText §5.3) or classifiers where robust (default) hyperparameter settings have been published (as in the case of BERT §5.1). We refrained from evaluating classifiers which are fairly sensitive to hyperparameter tuning and for which no generally applicable hyperparameter settings are known (as it is the case for biLSTMs).

5.1 BERT

In our experiments, we used the currently most sophisticated pre-trained model for BERT, namely BERT-Large, Cased: 24-layer, 1024-hidden, 16-heads, 340M parameters.[3] We experimented with two versions: one in which we fine-tune the model and an SVM [Joachims, 1999] that is trained on the BERT embeddings of the final layer.

We fine-tuned the BERT model on the respective training data by adding another layer on top of the existing model. The model was trained with standard hyperparameter settings: batch size: 32; learning rate: 2e-5; number of epochs: 3.

The embeddings for the SVM were obtained with the help of the script *extract_features.py* provided by Google research as part of BERT.[4] That script provides features (i.e. embeddings) from all layers of the pre-trained model. Following Reimers and Gurevych [2019], we only considered the final layer and averaged the embeddings representing the words of the sentence. Further details regarding the configuration of the SVM are presented in §5.2.

Since we did not measure any statistically significant difference between the two models, we decided in favor of the SVM since it is easier to handle in general. In addition, it allowed us to combine BERT with our linguistic features in a very simple and effective fashion.

5.2 SVM

In our experiments, we used SVM^{Light} [Joachims, 1999] as an implementation of an SVM. SVM is widely used for NLP-related tasks and is particularly effective on the detection of abusive language [Schmidt and Wiegand, 2017, Wiegand et al., 2018]. Since we have a dataset with a balanced class distribution, we felt that there was no need to tune specific parameters, particularly since in the default configuration we already obtained classification performance comparable to fine-tuned BERT (§5.1).

5.3 FastText

FastText [Joulin et al., 2017] has been promoted as a simple and efficient baseline system. Due to its model's simplicity, the tool is not dependent on extensive parameter-tuning. For all experiments the classifier was used with its *default configuration*. The only exception is that we also considered *pre-trained embeddings*. We experimented with the following publicly available models:

- Common Crawl (2 million vectors trained on 600 billion tokens) [Mikolov et al., 2018]
- Twitter (1.2 million vectors trained on 27 billion tokens) [Pennington et al., 2014]
- Wikipedia (1 million vectors trained 16 billion tokens) [Mikolov et al., 2018]

[3] `https://github.com/google-research/bert`
[4] `https://github.com/google-research/bert/blob/master/extract_features.py`

In our experiments, the embeddings induced from Common Crawl performed best. This can be explained by the fact that the model is trained on the largest text corpus.

A Situational Frames

1. (ILLNESS) You are at work, and a colleague comes in. He looks like he has a horrible cold. As you are talking, you want to express your concern at how ill he seems without offending him.
2. (STYLE) You are meeting a friend, who is trying out a new style. Unfortunately, the new style doesn't suit her. You are trying to comment on her appearance without offending her.
3. (EMOTION) You are meeting a friend, who arrives looking extremely upset, but doesn't talk about it. You are commenting on the fact that you have noticed how distraught your friend is, without offending him.
4. (TEAMWORK) You are in a meeting with a colleague who works on the same project as you. She hasn't done her part of the project well. You are trying to express your disappointment in her performance without offending her.
5. (EATING) You are having dinner with a friend, who has just come in from a long day, and is extremely hungry. You notice that his table manners are not the best and are trying to point this out without offending him.
6. (STRESS) You are a meeting a friend for lunch. She has been working long hours for the last few days, and when she arrives you can see that she is completely overworked. As you are talking, you want to express your concern at her appearance and her level of stress without offending her.
7. (HEALTH) It is a very hot day and you are outside with friends. You notice that one of these friends is particularly suffering from the heat. You are commenting on the fact that you have noticed this, without offending him.
8. (STRESS) You are at work, and you notice that the colleague next to you seems to be extremely tired today. You are commenting on her lack of sleep, without offending her.
9. (EMOTION) You are talking to a friend, who is due to give very important presentation very soon. He told you he wasn't nervous, but you can see signs of nervousity on him. You are trying to comment on how nervous he is, without offending him.
10. (DRIVING) You are getting a ride from a friend, who turns out to be a poor driver. You want to comment on her poor driving skills, without offending her.
11. (BEHAVIOR) You are at a bar, talking to a friend. He has made some comments which you find inappropriate about people you both know. You want to comment on how rude you find him, without offending him.

12. (BEHAVIOR) You are at a party with a friend, who hasn't been talking to the other guests much. You want to express that you think she is too shy, without offending her.
13. (HEALTH) It is a very cold day and you are outside with friends. You notice that one of these friends is particularly suffering from the cold. You are commenting on the fact that you have noticed this, without offending him.
14. (EMOTION) You meet up with a friend, who is obviously very sad about something. You want to comment on how sad she looks, without offending her.
15. (EMOTION) You have just told a friend some news that he found very shocking. You are commenting on how shocked he was, without offending him.
16. (BEHAVIOR) You have just overheard an argument between two of your friends, and now you are talking to one of them. You think that she behaved too aggressively during the argument. You want to comment on her words and behavior, without offending her.
17. (PERFORMANCE) You are at a party with a friend, and both of you are dancing. You notice that your friend is not a very good dancer. You want to comment on his dancing skills, without offending him.
18. (TEAMWORK) You have been working on a project with a colleague, who is extremely overenthusiastic and full of too much energy, which has sometimes been stressful for you. You want to comment on her behavior, without offending her.
19. (PERFORMANCE) You are at a party with a friend, who has unsuccessfully been trying to flirting with various people. He asks you for some advice and you try to point out what he's doing wrong, without offending him.
20. (BEHAVIOR) You just had a difficult conversation with a friend, and you are very disappointed in how she treated you. You are trying to point out some things she did wrong, without offending her.
21. (TEAMWORK) You have been working on a project with a colleague, who hasn't done his full share of the work, which has been very stressful to you. You want to comment on his behavior, without offending him.
22. (EMOTION) You meet up with a friend, who is obviously very upset and angry about something, but trying to hide it. You want to comment on how she looks, without offending her.
23. (STYLE) A friend has asked you for advice on his appearance and personal style. You are trying to point out negative things you have noticed about him, without hurting his feelings.
24. (DRIVING) You are getting a ride from a friend, who turns out to be a very reckless and fast driver. You want to comment on her driving skills, without offending her.

25. (STYLE) A friend has asked you for advice on her appearance and personal style. You are trying to point out things that she could change, without hurting her feelings.

26. (EATING) You are having dinner with a friend, who has cooked an extremely large amount of food, and given both himself and you big portions. You can't eat this much, and want to comment on it, without offending him.

27. (ILLNESS) You are at work, and a colleague comes in. He looks like he should have stayed home in bed that day. As you are talking, you want to express your concern at how ill he seems without offending him.

28. (BEHAVIOR) You are talking to a friend. She has made some jokes and comments that you didn't like. You want to comment on her words, without offending her.

29. (STYLE) You are at a friend's home, who is showing you his flat for the first time. As you walk through it, you notice many things you dislike, and try to comment on them politely and without offending him.

30. (BEHAVIOR) You are at a party with a friend, and she has been rejecting all attempts at flirting from different people in a harsh manner. You want to comment on her behavior, without offending her.

31. (PERFORMANCE) A friend of yours badly wants to be a successful musician, and has invited you to her show. It turns out that she is embarrassingly terrible. You are trying to tell her that you do not think she should be a singer, without hurting her feelings.

32. (ILLNESS) You are having lunch with a friend, who is obviously ill and barely eats anything of the food, but insists that he is fine. You want to express your concern at how ill he seems without offending him.

33. (PERFORMANCE) A friend of yours has been dating unsuccessfully, and asks you for advice on her appearance and behavior. You want to point out some negative traits you have noticed, without offending her.

34. (PERFORMANCE) You have just left a meeting together with a colleague of yours who is painfully shy. During the meeting, your colleague could not get his points across at all. You want to point out some problems with his approach, without hurting his feelings even more.

35. (PERFORMANCE) You are at a party with a friend, who has been trying to flirt with people unsuccessfully. She asks you for advice on her appearance and behavior. You want to point out some negative traits you have noticed, without offending her.

36. (DRIVING) You are getting a ride from a friend, who turns out to be a very reckless and fast driver. You want to comment on his driving skills, without offending him.

37. (HEALTH) You are meeting a friend whom you haven't seen in a while. When you see her, you are shocked by the changes in her and notice that something is obviously very wrong. You want to comment on her behavior and appearance, without hurting her feelings.

38. (TEAMWORK) You have been working on a project with a colleague, and the two of you don't get along well, which is starting to affect the work. You want to politely talk about some things that have been bothering you about him, without hurting his feelings.

39. (STYLE) A friend asks you for advice on her Halloween costume, which she made herself. You dislike the costume, and want to comment on her appearance, without hurting her feelings or insulting the work she put into it.

40. (EATING) You are having dinner with a friend, who has cooked for you. You notice that both his table manners and his cooking skills are not the best and are trying to point this out politely without hurting his feelings.

41. (HEALTH) You are meeting a friend whom you haven't seen in a while. She has been on an extreme diet. When you see her, you are shocked by the changes in her, and worried for her health. You want to comment on her behavior and appearance, without hurting her feelings.

42. (BEHAVIOR) You are talking to a friend. He has made some jokes and comments that you found hurtful. You want to comment on his words, without offending him.

43. (STYLE) You are at a friend's home, who is showing you it to you for the first time. As you walk through it, you notice many things that could be improved, and try to comment on them politely and without offending her.

44. (TEAMWORK) You have been working on a project with a colleague, and the two of you just had to give a presentation on it. You think your colleague isn't taking the whole thing seriously enough. You want to politely talk about some things that have been bothering you about the way he presents the project, without hurting his feelings.

45. (DRIVING) You are getting a ride from a friend, who turns out to be an extremely slow and overly cautious driver. You want to comment on her driving skills and the music in the car, without offending her.

References

Thorsten Joachims. Making Large-Scale SVM Learning Practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press, 1999.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of Tricks for Efficient Text Classification. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 427–431, Valencia, Spain, 2017.

Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. Advanced in Pre-Training Distributed Word Representations. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 52–55, Miyazaki, Japan, 2018.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. GloVe: Global Vectors for Word Representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Dohar, Qatar, 2014.

Ashequl Qadir, Ellen Riloff, and Marilyn A. Walker. Learning to Recognize Affective Polarity in Similes. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 190–200, Lisbon, Portugal, 2015.

Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China, 2019.

Anna Schmidt and Michael Wiegand. A Survey on Hate Speech Detection using Natural Language Processing. In *Proceedings of the EACL-Workshop on Natural Language Processing for Social Media (SocialNLP)*, pages 1–10, Valencia, Spain, 2017.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39 (2/3):164–210, 2005.

Michael Wiegand, Marc Schulder, and Josef Ruppenhofer. Separating Actor-View from Speaker-View Opinion Expressions using Linguistic Features. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 778–788, San Diego, CA, USA, 2016.

Michael Wiegand, Melanie Siegel, and Josef Ruppenhofer. Overview of the GermEval 2018 Shared Task on the Identification of Offensive Language. In *Proceedings of the GermEval Workshop*, pages 1–10, Vienna, Austria, 2018.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, pages 347–354, Vancouver, BC, Canada, 2005.