GERMS-AT: A Sexism/Misogyny Dataset of Forum Comments from an Austrian Online Newspaper

Brigitte Krenn, Johann Petrak, Marina Kubina, Christian Burger
Austrian Research Institute for Artificial Intelligence, DerStandard
Freyung 6 1010 Vienna Austria, Vordere Zollamtsstraße 13 1030 Vienna Austria
{brigitte.krenn, johann.petrak}@ofai.at, christian.burger@derstandard.at

Abstract

This paper presents a sexism/misogyny dataset extracted from comments of a large online forum of an Austrian newspaper. The comments are in Austrian German language, and in some cases interspersed with dialectal or English elements. We describe the data collection, the annotation guidelines and the annotation process resulting in a corpus of approximately 8 000 comments which were annotated with 5 levels of sexism/misogyny, ranging from 0 (not sexist/misogynist) to 4 (highly sexist/misogynist). The professional forum moderators (self-identified females and males) of the online newspaper were involved as experts in the creation of the annotation guidelines and the annotation of the user comments. In addition, we also describe first results of training transformer-based classification models for both binarized and original label classification of the corpus.

Keywords: sexism/misogyny dataset, sexism/misogyny classification, annotation guidelines

Content warning: We show illustrative examples of sexist and misogynous language to illustrate the annotation guidelines and to analyse error types.

1. Introduction and Motivation

The ever more widespread use of social media and user-contributed content also causes an increase of toxic or offensive language and other forms of unwanted contributions which may need to get detected and removed. In this paper, we present work aimed at supporting moderators of a large daily Austrian (German language) newspaper which allows registered users to discuss the articles published on its web-site. Users produce some 20K to 50K comments per day. An analysis of the commenting behaviour (carried out as an online user survey at the newspaper) has shown that only a third of the users participating in the online discussion are women and that one important reason why women avoid participating in article forum discussions is the presence of sexist/misogynist comments. This insight was a major motivation for creating the corpus in order to train classifiers which support moderators in detecting respective comments, so they can provide a more welcoming and safer climate of discussion especially for female users.

For this, a corpus of approximately 8000 comments was collected and annotated to be used for training classifiers to flag comments or entire discussion forums with a high number of suspected sexist/misogynist utterances. The dataset and a datasheet (Gebru et al., 2021) as well as the sample comments provided to the annotators for training will be made available from https://huggingface.co/datasets/ofai/GerMS-AT. The dataset is distributed under CC BY-NC-SA 4.0.

As for the definition of sexism and misogyny, we rely on the definition given in Encyclopedia Britannica which defines **sexism** as "prejudice or discrimination based on sex or gender, especially against women and girls", and which defines **misogyny** as follows: "The extreme form of sexist ideology is misogyny, the hatred of women."[1] These definitions are in line with the definitions used in the SemEval 2023 Task 10 Explainable Detection of Online Sexism (Kirk et al., 2023), p. 2194: "Misogyny refers to "expressions of hate towards women" (Ussher, 2016), while sexism also covers more subtle implicit forms of abuse and prejudice that can still substantially harm women." As the presented dataset comprises comments which are either sexist or misogynist or both, we use *sexism* or *sexist* to refer to sexist or misogynous comments in the reminder of the paper.

Contributions of our work: To the best of our knowledge, we present the first German dataset of *sexist* forum comments annotated with labels ranging from 0 (not *sexist*) to 4 (extremely *sexist*), whereby 67 % of the assigned labels were 0 and 33 % ranged from 1 (mildly *sexists*) to 4. Individual comments are annotated by 2 (1400 comments), 3 (6496) and 8 (99) annotators.

2. Related Work

Together with work on toxic and offensive language classification in recent years, there has

[1] Both quotes are from https://www.britannica.com/topic/sexism.

LREC-COLING 2024, pages 7728–7739
20-25 May, 2024.

also been increasing work on the classification of sexist or misogynist language, sometimes as part of a more general toxic language classification task. Hewitt et al. (2016) give an overview over earlier work and describe a dataset of English Tweets containing abusive sexist terms. Anzovino et al. (2018) present work on creating a dataset of tweets, subsequently used as part of IberEval-2018 and EvalIta-2018 challenges (Fersini et al., 2018) for sexism classification where English and Spanish datasets were made available. Shushkevich and Cardiff (2019) give an overview over misogyny detection in social media, specifically Twitter. Waseem and Hovy (2016) describe work on an English corpus of 16K tweets for detecting toxic and hate speech including sexist slurs or defending sexism (3383 tweets with sexist content). Frenda et al. (2019) present work on the English and Italian language datasets described in (Fersini et al., 2018) and (Waseem and Hovy, 2016). Sharifirad and Matwin (2019) include some more detailed description of sexist language and is based on another dataset of English language tweets. Parikh et al. (2019) describe work on categorizing accounts of sexism from the Everyday Sexism Project website through fine-grained multilabel classification. Other datasets containing sexism are described or used in: Chiril et al. (2020a), 12K tweets in French; Grosz and Conde-Cespedes (2020), tweets, work-related quotes, press quotes and other sources; Bhattacharya et al. (2020) and Safi Samghabadi et al. (2020), Youtube comments in Indian English, Hindi and Bengla; Rodríguez-Sánchez et al. (2020), Spanish language tweets; Mulki and Ghanem (2021), 6603 Arabic tweet replies scraped from the time-lines of popular female journalists/reporters during October 17th protests in Lebanon; Zeinert et al. (2021), Danish language dataset sampled from several social media sites; Jiang et al. (2022), 8969 Chinese comments from Sina Weibo, Figure 1 of their paper also provides a list of sexism-related datasets for various languages. The EXIST task at IberLEF 2021 (Rodríguez-Sáchez et al., 2021) addresses the identification and categorization of sexism in English and Spanish language tweets and comments from `Gab.com`. Most recently, SemEval 2023 Task 10 (Kirk et al., 2023) describes an English language corpus of 20000 texts sampled from Gab and Reddit, annotated with 3 hierarchical labels. The paper also presents the top ranking systems for the different tasks: Task A binary classification (sexist versus non-sexist); Task B identification of four distinct categories of sexism, comprising 1) threats, plans to harm and incitement, 2) derogation, 3) animosity, and 4) prejudiced discussion; and a more complex Task C which further breaks up the four sexism categories into more fine-grained classes (11 in total, with two up to 4 classes per category).

German datasets comprising sexism-related labels are scarce: The Moderator- and Crowd-Annotated German News Comment Datasets (Assenmacher et al., 2021) contains labels for sexism, racism, threats, insults, and profane language, however, only 1530 comments out of 85K are annotated as sexist by at least one annotator. The DeTox dataset for German Offensive Language includes amongst others sexual identity as type of discrimination which, however, applies only to 73 out of over 10K comments (Demus et al., 2022). In comparison, our dataset comprises 3596 comments annotated as *sexist* out of 7995 comments in total.

Sexism or misogyny datasets are usually annotated with hierarchical labels of sexism categories. This differs from our approach, where we define *sexism* categories in the annotation guidelines to provide a solid basis for the annotators to decide whether a comment is *sexist* or not, whereas the *sexist* comments are then labelled with (subjective) degrees of severity of the *sexism* expressed.

3. Corpus Creation

3.1. Data Collection

Comments to be annotated were collected from several different sources (The comments most likely to contain *sexist* data stem from 2020 to 2022, comments that most likely contain non *sexist* contributions stem from a larger time span.):

1. a collection of comments which had been reported by forum users with a (free text) reporting reason that included a keyword related to sexism or misogyny,

2. comments which were reported with a different reason,

3. comments randomly sampled from all available comments,

4. a subset of comments (in 2.) pre-classified with an early version of the binary classifier trained on the first 2800 comments annotated with labels 1 for *sexist* and 0 for not *sexist*,

5. comments from 24 article forums which were manually identified by forum moderators to contain an above-average number of comments considered *sexist*. These comments were pre-classified with the same early binary classifier as used in 4. From these pre-classified comments, two sets were selected to be part of the dataset prepared for annotation: The one set comprises the highest probability label 1 comments in order to correct

false positives. The other set comprises those label 0 and 1 comments with close to 0.5 probability in order to add what may be hard to classify instances.

Comments in the dataset are between 3 and 999 characters long, with a median of 145 characters and an interquartile range of 216 characters. See Table 1 for an overview of the key figures. Note: The selection of comments did not involve any constraints on comment length apart from the requirement that all comments had to have a minimal length of 3 characters.

mean	213.72	25%	73
std	194.44	50% (median)	145
min	3	75%	289
max	999		

Table 1: Distribution of comments in number of characters.

3.2. Data Cleaning and Anonymization

The comment texts are present in the dataset with the original newline and whitespace characters preserved. For anonymization, the following changes were made to the text: URLs were replaced with the placeholder `{URL}`. At-mentions (e.g. `@name`) were replaced with `{USER}` (170 occurrences). Any email addresses would have been replaced by `{EMAIL}` but none were present in the text. To further protect the privacy of individual commenters, each comment was manually checked for potential mentions of user names or nick names. This was done in an additional annotation round by a single annotator. The thus identified user names were then systematically replaced with the placeholder `{USER}` (149 occurrences). In total, 179 comments contain one or more placeholders. Documents of length < 3 characters and documents only containing a URL were removed. Further means for privacy protection were: only the plain comment text was kept, and deleted were (i) all meta information regarding user names and nick names, (ii) all information indicating the position of a comment within a certain thread, (iii) all information which would allow a comment to be associated with a particular forum. This means have clear effects on the annotation, as decisions whether a comment is *sexist* or not and to what extent must be made on the basis of the individual comment text only, without the availability of further context.

3.3. Choice of Annotators

The aim of the manual corpus annotation was to reflect the judgement of moderators in their everyday work. For this reason, the manual annotations were carried out by up to 8 annotators of which 7 were experienced moderators. The 8th annotator was a natural language processing and corpus linguistics expert. Two of the annotators are among the authors of this paper. There were 3 annotators who self-identify as male (all working as forum moderators at the newspaper) and 5 who self-identify as female (4 of which working as moderators at the newspaper). All annotators are native speakers of German.

3.4. Annotation Paradigm and Definition of Label Set

Annotation paradigm While (Röttger et al., 2021) argue for a decision between the descriptive and prescriptive annotation paradigm when annotating a dataset, we explicitly aimed for a combination of both: (i) We pursued a prescriptive approach with the list of criteria determining what should be classified as *sexist* as described in the annotation guidelines (Section 4). These criteria cover the newspaper's gender policy and forum netiquette. (ii) We pursued a descriptive approach by asking the annotators (all but one experienced forum moderators at the newspaper) to annotate those comments they have classified as *sexist* on a scale from 1 to 4 according to their personal perception of the severity of sexism expressed in the comment. The annotators were instructed to ask themselves "How uncomfortable do I feel reading this comment?". Asking annotators for a graduated, subjective assessment about the severity of a user comment considered as sexist allowed us to create a dataset which captures gradations in the assessment of sexist utterances even within a rather homogeneous group of professional content moderators with experience of moderating sexist comments against women. In pursuing both a prescriptive and a descriptive approach to annotation, we on the one hand created a dataset for the training of classifiers which support the moderation goal to create a welcoming atmosphere for female contributions to forum discussions. In particular, we were able to replicate the moderators' manual assessment of forums (not part of the dataset) as containing more sexism versus little to no sexism applying the by then best classifier model trained with the GERMS-AT corpus (Petrak and Krenn, 2022). Apart from its use for binary classification (*sexist* versus *non-sexist*), this corpus on the other hand is aimed for use in machine learning research into how to make models aware of more or less disagreement on the label, i.e., understand subjectivity versus the up to date prevalent approach to corpus annotation assuming a ground truth where diverging annotator opinions need to be unified (majority vote, subsequent consensus by the annotators, or decision by a meta reviewer).

Label set Comments were annotated by assigning one of 5 possible labels $0 \ldots 4$, corresponding to 0 = absence of *sexism* and 4 levels of "severity" of the expressed *sexism* as perceived by the individual annotators, with 1 = mild, 2 = present, 3 = strong, 4 = extreme.

3.5. Annotation Process

Comments were given to the annotators in batches of 100, by creating a spreadsheet from the comment texts and preparing a selection field for selecting one of the 5 possible labels, see Figure 1. The first batch of 100 comments was given to all 8 annotators and then analysed to find comments with the biggest disagreement. For this we calculated a heuristic disagreement score based on all pairwise distances between the labels, where the distance between labels 0 and 1 were defined to be 4, and distances between labels $l_i, l_j > 0$ were defined to be $|l_i - l_j|$. Examples with high scores were then discussed among annotators to revise the annotation guidelines and clarify misunderstandings.

Thereafter, each round of 100 comments was given to a random selection of 3 annotators, making sure that everyone is paired with everyone else from the annotator pool equally often. This was done in order to compromise between annotating as many comments as possible given the available time and resources and still get enough annotators for each comment to identify disagreements. In total, 6496 comments annotated by 3 annotators, 1400 by 2 annotators and 99 by 8 annotators remained in the final dataset[2], resulting in a total of 23080 annotations. The overall distribution of assigned labels is shown in Table 2.

Label	0	1	2	3	4	$1 \ldots 4$
%	67.0	8.1	13.8	8.7	2.4	33.0

Table 2: Distribution of assigned labels based on all annotations.

4. Annotation Guidelines

Since the phenomenon of sexism and misogyny is complex, guidelines were created to describe the most important kinds of *sexism* relevant for the annotation task. For this we built upon the newspaper's netiquette and forum policy regarding *sexist* user contributions, and the moderators' experiences with and how to curtail *sexism* in forum discussions. In addition, we used existing categorizations (e.g., (Parikh et al., 2019)[3] as further input to the guideline development. The guidelines also attempt to clarify some of the difficulties likely to be encountered: how to decide if there is not enough context, what if the sexist remark is aimed at a man or men in general, how to treat "reported sexism" (Chiril et al., 2020b). Above all, the guidelines follow the aim of providing help for annotating in a way that reflects the daily work of moderators and the newspaper's editorial concept. They are not meant as an accurate abstract definition of sexism and misogyny, but as a means to delineate the frame under which a user comment should be considered as sexist. The labels assigned by the annotators reflect that one or more of the defining criteria is present, and to which extent (1 mildly sexist ... 4 extremely sexist). Any combination of criteria is sufficient for a positive instance, and the scores are subjective.

4.1. General Instructions to the Annotators

As described in Section 3.2, each comment was to be annotated as standalone utterance, without considering the larger context of the thread or the particular forum it occurs in. This means that a comment is manually annotated as *sexist* when it has a *sexist* reading without further context. However, the forum moderators due to their moderation experience have a personal assessment whether a comment is *sexist* or not. Accordingly, the moderators were instructed to take the following into account when annotating a comment:

- When reading a comment, a *sexist* context should not be constructed away if it is rather obvious from one's own moderation experience.
- Equally, when reading a comment, a *sexist* context should not be fantasized about. One should be particularly critical of one's own potential bias, especially if one has been annotating several *sexist* comments in a row. The following three examples for comments and related annotation decisions were given to the annotators.

Examples:

- comment: "Bei aller Tragik und Ernsthaftigkeit … wir haben schon a fesche Justizministerin" (EN: "With all tragedy and seriousness ... we

[2] 5 comments with their annotations were removed after all annotations were completed, due to the decision to delete comments containing only a URL and comments < 3 characters.

[3] Parikh et al. is an example for a more extensive collection of categories, i.e., 23 categories of sexism which were formulated, as stated in their paper, under the direction of a social scientist taking into account gender-related discourse and campaigns.

	Label	Posting	PID
2	0-Kein	Als ob man mit einer Wand redet. Schließens nicht von Müttern auf Frauen. Mütter sind ein kleiner Teil der Frauen. Zusätzlich zum Beruf Hausarbeit erledigen muss also kein Mann machen? Meine Wäsche wäscht sich also laut Ihnen allein und meine Wohnung putzt sich von selbst oder ich mache das in meiner Arbeitszeit oder woher kommt dieser Unsinn?	1067876521
3		Wenn der erste schwarze Transmann den Mond in einem Rollstuhl befährt hab ich einen Grund zu feiern. Alles andere sind müde Beschwichtigungen von alten weißen Astronauten!	1067804672
4	0-Kein 1-Gering 2-Vorhanden 3-Stark 4-Extrem	Haha. Ich stell mir gerade vor, wie der Palantir-CEO nach einem entspannten Meeting mit dem CIA-Chef seine Mails checkt und vor Ehrfurcht zittert, weil er eine Mail vom Werner Faymann kriegt. Da muss er die Rudas natürlich gleich zum Vice President Strategy machen, wenn auch noch der Häupl und die Dorli Bures Druck machen.	1055431144
5		Frauen werden gegeneinander ausgespielt? Mit Erfolg? Ich dachte, Frauen sind sozialer?	1057846767
6		Nun die hat halt etwas länger gebraucht und keiner aus ihrem Freundeskreis bis nach der mittelschule würde zu ihr gehen. Und wie gesagt mit etwas Kohle und guten Beziehungen ist so manches einfacher.	1055450233
7		Der Hr. Norbert Hofer hat sie in den Aufsichtsrat der Austro-Control gebracht. Qualifiziert wodurch? Spendierfreudig?	1065973394
		danke für den freundlichen und wertschätzenden ton. ich hätte es auch nicht als sarkastisch empfunden. keine sorge, der ärger stammt nicht von diskussionen in online-basierten diskussionsräumen, sondern rührt daher, dass wir in 2021 echt schon weiter sein sollten. diesbezüglich darf man gar nicht un-aufgeregt sein, finde ich. feminismus wird leider immer noch viel zu oft (und zu reflexartig) als „angriff gegen männer" verstanden (der ganze diskussionsthread hier ist ja ein sehr schönes beispiel dafür). ganz ehrlich: einen gender care gap mit knapp 1,5 stunden pro tag rede ich mir ja nicht ein.	1067883824

Figure 1: The annotation sheet providing per line the drop-down menu for selecting the label (kein – not *sexist*, gering – mild *sexism* , vorhanden – present, stark – strong, extrem – extreme), the comment text to be annotated, and a unique ID per comment (PID).

definitely have a dashing Minister of Justice") – Annotation decision: ***sexist*** because the appearance of the Minister of Justice plays no role in the reporting about her as Minister of Justice.

- comment: "sie müssen echt hübsch sein…" (EN: "they must be really pretty...") – Annotation decision: **not *sexist*** because without context it is not 100 % clear that "sie" refers to females, and if it does one cannot know if it is meant pejoratively (and therefore sexist) or not.

- comment: "Einfach tief von unten hineinfahren und Batterien rausnehmen" (EN: "Just dive in deep from below and take out batteries") – Annotation decision: **not *sexist***, however, it is a **difficult case** as a sexual act may be referred to, but without context it is not ascertainable.

4.2. Criteria to Identify a Comment as *Sexist*

In the following, the criteria according to which a comment is considered *sexist* are presented. Per criterion an example is given, referred to as (Ex. A.x). The respective examples can be found in Appendix A (content warning: *sexist* comments!):

4.2.1. Generalizing stereotypes - attributions to groups of women

Role stereotypes (Ex. A.1.1): generalizations about certain roles that are better suited for women, such as: (i) Women are better suited for housework, parenting, social jobs, etc. (ii) A woman should have long hair, wear skirts, and be made up.

Attribute stereotypes (Ex. A.1.2): linking women to some physical, psychological behavioral qualities or likes/dislikes such as (i) Women always feel offended/women put themselves in the victim role. (ii) Women can not think logically and have no place in science. (iii) Women are too weak for certain jobs. (iv) Women choose men who are successful and earn good money. (v) Women spend the men's money. (vi) Devaluation of supposedly feminine qualities such as being soulful, caring, sensitive, etc.

4.2.2. Reduction of a person to her appearance (Ex. A.2)

(i) The appearance is praised or evaluated or addressed as something fundamentally necessary for being a woman; (ii) the appearance is put in relation to something, e.g., to the performance; (iii) questioning a person's femininity; (iv) body and fat shaming.

4.2.3. Women as sexual objects (Ex. A.3)

(i) Statements about a person's appearance that sexualize; (ii) sexually charged terms; (iii) salacious comments about a woman, either named in the article the forum is related to or towards a female commenter.

4.2.4. Female connoted insult (Ex. A.4)

Certain terms, insults that have *sexist* connotations.

4.2.5. Denigration of women, their performance and women's issues (Ex. A.5)

Denial of female performance, denial of the existence of gender differences in salary, all kinds of female attribution: (i) Denying that a woman got the job because of her qualifications by (a) calling the woman a token woman, or (b) claiming she slept her way up. (ii) Claiming that women choose the "wrong" education/fields of study and therefore are not in leadership positions/are not successful/earn less than men. (iii) Trivializing of problems specific to women or of structural inequalities by (a) dismissing inequalities and structural reasons for the pay gap referring to the part-time rate; (b) claiming that women voluntarily work part-time, so it is their own fault that they earn less; (c) negating care work, much of which is done by women. (iv) Disparagingly opposing gendering by (a) using malapropism; (b) dismissing the effect of gendering with abstruse arguments that women do not want a gender-sensitive language anyway.

4.2.6. Downplay sexual violence and sexual harassment against women (Ex. A.6)

(i) Perpetrator-victim reversal; (ii) presenting #Metoo as nonsense in an unobjective way; (iii) as a man, judge what falls under sexual harassment against women; (iv) the woman uses allegations of sexual harassment for a purpose that serves her.

4.2.7. Whataboutism (Ex. A.7)

Claiming that men are much more likely to be affected by violence, women do not work in heavy labourer's jobs such as in construction, garbage collection, etc.

4.2.8. Abortion (Ex. A.8)

(i) Abortion is equated with murder. The woman is thus accused of criminal behavior that would presumably not be attributed to a man, since he does not carry the child to term. (ii) The woman's self-determination is questioned or denied.

4.2.9. Misandry (Ex. A.9)

Given a *sexist* utterance against men, can the male referent be replaced by a female referent and does the resulting utterance fall under one of the above categories? If yes, the utterance is assigned a *sexism* label from $1 \ldots 4$. For instance, if "tote Männer" (dead men) in Ex. A.9 is replaced by dead women it would be sexist against women, because the comment generalises over a whole group of individuals.

5. Annotator Agreement and Corpus Analysis

Krippendorff Alpha over all annotations was 0.38 (nominal scale) and 0.64 (ordinal scale). After binarization of the 5 possible annotations into 0 (for no sexism) and 1 (for labels $1 \ldots 4$), Krippendorff Alpha was 0.60. Overall agreement was 0.65 / 0.83 (binary) if macro averaged over all agreements of pairs of annotators, and 0.67 / 0.83 (binary) if micro averaged over all pairs of annotations. F1.0 macro over all 23660 pairs of annotations was 0.40 / 0.81 (binary). Overall Cohen's Kappa was 0.39 / 0.64 (binary) if macro averaged over all kappas for pairs of annotators. This indicates that there was considerable difference of opinion or difficulty in assigning the fine-grained labels.

As shown in Table 2 the overall rate of assigned positive labels (classes $1 \ldots 4$) was 0.33. Looking at all pairs of annotations in the dataset the relative frequencies of annotation pairs (confusion matrix) is shown in Table 3. This illustrates the large rate of disagreement among annotators, especially on estimating the fine-grained degree of *sexism* (labels $1 \ldots 4$).

	0	1	2	3	4
0	0.524	0.032	0.036	0.015	0.002
1	0.032	0.014	0.020	0.009	0.001
2	0.036	0.020	0.052	0.037	0.007
3	0.015	0.009	0.037	0.045	0.017
4	0.002	0.001	0.007	0.017	0.014

Table 3: Relative frequency of annotation pairings.

We deliberately did not decide on a final single "best" label as the judgement on *sexism* may depend on personal opinion and we believe it would be wrong to assume there is a single correct value for each instance. Other instances may have received labels by mistake (e.g. misinterpretation of the text or of the annotation guidelines). For creating a training set we used different strategies to resolve disagreements (see Section 6).

To the best of our knowledge, this is the first German language corpus related to *sexism*. The corpus also differs from most other sexism corpora in that it originates from the comments of a site which is very strictly moderated and where sexism/misogyny is present often in very subtle ways or phrased ambiguously which can cause disagreement between annotators about the presence and severity of the sexism in the comment.

5.1. Qualitative Analysis of Annotator Disagreements

In the following, we discuss the 100 strongest disagreements between 3 annotators. Recall, the majority of comments was annotated by 3 annotators, leading to 3 scores from 0 to 4. From these 3 numbers the mode, the median and a binary number (representing the majority vote *sexist* (+ coded as 1, if two of the three annotators annotated a label from $1 \ldots 4$) versus non *sexist* (- coded as 0, if two of the three annotators annotated 0). The binary value provides us with the majority vote, mode and median tell us about the central tendency, whereby mode is not always defined, i.e., in those cases where each of the 3 annotators gave a different number. As we already see from Table 4, the majority of the top-ranking discrepancies is when two of 3 annotators judged the comment as *sexist* and one as not *sexist*. For instance in 20 cases, one annotator judged a comment as extremely *sexist* (4), one as strongly (3) and the third one as not *sexist* (0). In 23 cases, the verdict was extremely (4) versus mildly (2) versus not (0) *sexist*. Another frequent case (18 out of 100) is when two annotators judge a comment as not (0) *sexist* whereas another one judges the comment as extremely (4) *sexist*.

#	labels	mode/median	binary
18	$4\ 0\ 0$	0/0	-
9	$4\ 1\ 0$	NA/1	+
23	$4\ 2\ 0$	NA/2	+
20	$4\ 3\ 0$	NA/3	+
4	$4\ 4\ 0$	4/4	+
9	$3\ 0\ 0$	0/0	-
14	$3\ 2\ 0$	NA/2	+
3	$3\ 3\ 0$	3/3	+

Table 4: The 100 most diverse annotations: distribution of label combinations (number # of instances with label combination), measures of central tendency (mode/median) and binary label.

From a manual inspection of all 100 comments, we found that a larger portion of the ambiguous cases is related to (i) sexualization such as female as sexual actor or object, and relation to prostitution; (ii) belittling or devaluation including: masculinization of women, reference to their ugly looks, spiteful comments, the looks of the person show their negative character, missing sexual attraction, devaluation because of the person's sexual orientation, missing skills or intelligence, intelligence as negative attribute, social climbing via sexual intercourse, prostitution, women should go back to the stove, and home work is no work.

6. Classification Models

The main purpose of creating a deployed classification model was to alert moderators both of individual *sexist* comments and article forums with a high rate of potentially *sexist* comments. For this reason, our main interest is a binary classifier (original label 0 vs original labels $1 \ldots 4$). However, we also studied the performance of a model for predicting the original label, both as seen as a multiclass classification task and as an ordinal regression task. Finally, we investigated if combining both the binary and multiclass tasks into a multi-task model would impair or improve the performance of the individual tasks.

All models are based on a transformer architecture (Vaswani et al., 2017) with one or two classification heads on top of the pooling layer. We used the pre-trained German BERT models `gbert-base`[4] and `gbert-large`[5] (Chan et al., 2020). For all models, accuracy and F1.0-macro metrics where estimated using 5-fold cross-validation with class-stratified folds. 10% of the training set were used as dev-set and this split was class-stratified as well. Hyper-parameters were explored in a step-wise process based on the first 6000 annotated examples as they became available. For this, manually chosen values for a few hyper-parameters were evaluated using 5-fold cross-validation and grid search. For many parameters, however, there was no clear best selection as the range of F1.0-macro estimation results introduced merely by choosing different random seeds was larger or comparable to the changes in F1.0-macro estimation for different parameter values. In such cases we chose an intermediate value among those with similar estimates. The final set of hyper-parameters used for all models described below was mostly identical to the original default values: BERT maximum sequence length 192, batch size 8, gradient accumulation: 1 batch, learning rate 7.5e-06, language model dropout rate 0.1 and no layer-wise learning rate adjustment. The AdamW optimizer with weight decay 0.01 and linear warm-up during 200 training steps was used. For all classification heads we used an additional layer with 768 hidden units, ReLU and no dropout before the actual output layer.

We evaluated the following single-head models: Bin (binary classification), Multi (multiclass classification), Coral (multiclass classification using an implementation of the CORAL ordinal regression model (Cao et al., 2020)); and the following dual-head models: BinMulti (binary and multiclass heads combined), BinCoral (binary and CORAL heads combined). For each of these 5 models we evaluated a variant based on the `gbert-base` model (/B) and one based on the `gbert-large` model (/L).

Table 5 shows the estimation results (accuracy and F1.0-macro) on a training set where the original and binarized targets where selected as the most frequently assigned labels, falling back to the highest most frequently assigned label or the maximum assigned label. For dual-head multi-task models there are two lines, showing the binary model as head 1 (:1) and the multiclass model as head 2 (:2). For the binary classifier, the dual-head binary classifier in combination with the Coral ordinal regression head, based on `gbert-large` achieved the highest accuracy (0.767) and the

[4] `https://huggingface.co/deepset/gbert-base`
[5] `https://huggingface.co/deepset/gbert-large`

best F1.0-macro result (0.741). For the original multi-class data the same model also achieved the (0.663), however the highest F1.0-macro result was achieved by the multiclass classifier from the dual-head classifier in combination with the binary classifier, based on `gbert-large` (0.330). As can be seen `gbert-large` based models consistently achieved better results than models based on `gbert-base` with regard to accuracy, and most of the time with regard to F1.0-macro estimates. There is some indication that multi-head/multi-class training can benefit the quality of at least one of the classification heads.

Model	Accuracy	F1.0 macro
Bin/B	0.738±0.005	0.712±0.014
Bin/L	0.748±0.050	0.674±0.154
Multi/B	0.578±0.045	0.320±0.016
Multi/L	0.633±0.026	0.313±0.087
Coral/B	0.634±0.022	0.260±0.016
Coral/L	0.657±0.014	0.261±0.055
BinMulti/B:1	0.729±0.017	0.705±0.015
BinMulti/B:2	0.576±0.027	0.310±0.017
BinMulti/L:1	0.759±0.020	0.733±0.017
BinMulti/L:2	0.607±0.025	**0.330**±0.021
BinCoral/B:1	0.735±0.018	0.707±0.012
BinCoral/B:2	0.652±0.007	0.256±0.021
BinCoral/L:1	**0.767**±0.014	**0.741**±0.015
BinCoral/L:2	**0.663**±0.010	0.275±0.023

Table 5: Accuracy and F1.0 macro estimates (± standard deviation) for models trained on the most frequent original (coral, multi) and binarized (bin) sexism label.

Model	Accuracy	F1.0 macro
Bin/B	0.721±0.011	0.720±0.012
Bin/L	0.650±0.137	0.589±0.214
Multi/B	0.506±0.007	**0.362**±0.008
Multi/L	0.516±0.025	0.333±0.104
Coral/B	0.548±0.011	0.281±0.027
Coral/L	**0.559**±0.015	0.264±0.069
BinMulti/B:1	**0.730**±0.011	**0.728**±0.012
BinMulti/B:2	0.498±0.005	0.336±0.023
BinMulti/L:1	0.713±0.093	0.672±0.178
BinMulti/L:2	0.522±0.040	0.317±0.098
BinCoral/B:1	0.723±0.003	0.721±0.002
BinCoral/B:2	0.541±0.008	0.271±0.011
BinCoral/L:1	0.713±0.092	0.674±0.178
BinCoral/L:2	0.558±0.017	0.269±0.074

Table 6: Accuracy and F1.0 macro estimates (± standard deviation) for models trained on the highest original (coral, multi) and binarized (bin) sexism label.

We also trained the same set of models on a training corpus where both the binary and multi-class target was always selected as the highest label assigned by any annotator ("when in doubt, treat it as *sexist/mysogynistic*"). The results for this experiment are shown in Table 6. On this data, the best accuracy (0.730) and F1.0-macro result (0.728) for the binary problem was achieved by the binary classification head of the dual-head model in combination with the multiclass-head, based on the `gbert-base` model, the best multi-class accuracy (0.559) was achived by the single-head Coral ordinal regression model, based on `gbert-large` while the best multiclass F1.0-macro result was achived by the single head multi-class model based on `gbert-base`. Some of the variants achieved very similar results (e.g. Coral/L and Bin-Coral/L:2). On this problem there is also less indication of an improvement of `gbert-large`-based over `gbert-base` based models or dual-head models over single-head models.
The code for all experiments is based on the FARM library[6] and is available online[7].

7. Conclusion

We have presented a dataset of approximately 8000 user comments from an Austrian German online newspaper. Apart from labelling the data on a binary level (*sexist* yes/no), the *sexist* class is labelled with a subjective evaluation of the severity of the *sexism* expressed in the comment. Because of the multiple annotations (2 to 8 annotators per comment), the corpus is also a valuable resource for the investigation of how to deal with disagreement and modeling diversity of opinion in machine learning. The corpus contains the text of the comments (with no information to identify the author) as well as all original labels (with no information about the annotator except their self-identified gender), and the identification of the source of the comment (cf. Section 3.1).
The data was collected for creating a deployed classification model to alert moderators of individual *sexist* comments as well as of forums with a high proportion of *sexist* comments. First experiments were conducted with binary classification models and models for predicting the original label seen as multiclass classification or ordinal regression tasks. The German BERT models `gbert-base` and `gbert-large` and the CORAL ordinal regression model realized as single and dual head models were tested. The results obtained in those first experiments reflect the difficulty of identifying subtle or ambiguous sexism/misogyny in text and the F1.0 results are only a few percentage points below the F1.0 metric calculated

[6]https://github.com/deepset-ai/FARM
[7]https://github.com/OFAI/paper-lrec2024-code

as inter-annotator agreement over the manual annotations.
Unlike most corpora about sexism and mysogyny in other languages, this corpus does not focus on mysogyny and sexism in the form of offensive comments or strong language but rather contains predominantly texts with subtle, ambiguous and veiled forms of sexism on which human annotators often disagree. As such, we believe it contributes a valuable dataset for more research into how to adapt machine learning models to these properties of a corpus.
We publicize the dataset with the hope that it will be used not only for replicating and extending the work on creating useful classification models from it, but also to update or add to the labels or extend the corpus with additional texts in order to better understand how people may disagree on what should be considered to be sexism and misogyny. For this reason, the dataset is made available as a git repository which is meant to receive updates, fixes or modified version by contributors and see new releases in the future.

8. Acknowledgements

This work was conducted as part of the projects FemDwell[8] supported through FemPower IKT 2018[9] and EKIP – A Platform for Ethical AI Application[10] supported by the Austrian Research Promotion Agency (FFG)[11]. We thank the forum moderators at derStandard.at for their contributions to developing the annotation guidelines and their efforts in annotating the dataset. Last but not least, we are grateful for the invaluable comments of the three anonymous reviewers.

9. Ethical Considerations and Limitations

We have presented a dataset of sexist and misogynous comments collected from forum comments of an Austrian German language online newspaper. The dataset was collected with the aim of training classifiers that support the forum moderators to monitor individual fora with regards to arising sexist and misogynous discussions. According to the moderators, it is crucial that harmful comments are identified early on in order to bring back a discussion to a respectful tone. 20K to 50K comments with rising tendency are made per day in the newspaper's forum, which is impossible to manually monitor. Therefore, automatic monitoring with the help of classifiers is a precondition for moderators to be able to intervene already at the upcoming of a harmful and intimidating discourse.

Creating, maintaining and publicizing a respective training corpus requires ethical consideration regarding affected individuals and groups when collecting, cleaning and annotating the dataset, when developing classifiers based on the dataset, when deploying classifiers trained on the dataset, when making publicly available the dataset and the trained models and/or the code for model training. The **annotators** of the dataset risk to be harmed by repeated exposure to sexist and misogynist utterances. Even though almost all annotators are professional forum moderators and used to handling sexist and misogynous comments, regular monitoring is called for to watch for negative effects of excessive exposure to harmful content on individuals. To mitigate harmful effects, the comments to be annotated were distributed in patches of 100, sampled from a mix of sources more or less likely to contain *sexist* utterances, the same batch was given to at least 2 annotators, the annotation process was organised in annotation rounds with several patches and all annotators of one round took part in regular check-ins with the whole team. These and later check-ins during the model development process also included the **researchers and developers** of the classifier models. Especially those who are involved in the qualitative analysis of the model outcome might be particularly affected by the exposure to harmful content. Likewise, **readers** of the paper may be negatively affected by the *sexist* content. The exposure to sexist and misogynous content may also lead to prejudiced discussions and the reproduction or reinforcement of harmful representation stereotypes. To raise awareness of harmful content, content warnings are placed well before examples for *sexist* comments are presented in the paper, cf. (Kirk et al., 2022). **Forum users** whose comments are part of the dataset and forum users who might be mentioned in the comments risk violation of their privacy. As countermeasures only the comment texts (without reference to a forum or a thread) are used and all potential user names, at-mentions, URLs, email addresses are deleted. **After deployment** of the trained model, **forum moderators** need to be schooled how to interpret the model outcome, and their awareness must be raised that the model decision is subordinate to their human expert decision. **Forum users**, on the one hand, benefit from the improved discussion climate in the forum. On the other hand, some of them may be puzzled because a comment was deleted, and respective explanations should be given. Therefore as part of the model deployment, it is essential that a concept and its technical realization for informing forum users about moder-

[8] https://ofai.github.io/femdwell/
[9] https://austrianstartups.com/event/call-fempower-ikt-2018
[10] https://ekip.ai/
[11] https://www.ffg.at/en

ation decisions is implemented. The **public availability of the dataset** allows other researchers to take up and further extend the work. To better understand the dataset, its capacity and limitations, a datasheet is published together with the dataset. Likewise, we strongly recommend to publish a model card (Mitchell et al., 2019) with each model trained on the dataset. Despite the precautionary measures indicated above, misuse cannot be completely ruled out, including the risk of improper use of the dataset in training sexist bots.

Limitations comments were annotated without further context, thus, actually *sexist* comments may be missed where the larger context (be it the article a forum is related to, or the thread a comment is part of) is decisive for the interpretation of their content. The annotations are geared towards a particular newspaper's netiquette and forum moderation policy. In other contexts, other criteria of what is considered *sexist* may hold and what is the level or severity of the *sexism* expressed. Moreover, what counts as sexist or misogynous and its expression is likely to change over time, which requires the dataset as well as the classifiers to be regularly updated. This, we hope to achieve by making the dataset publicly available and by supporting proper versioning.

10. Bibliographical References

Maria Anzovino, Elisabetta Fersini, and Paolo Rosso. 2018. Automatic identification and classification of misogynistic language on twitter. In *International Conference on Applications of Natural Language to Information Systems*, pages 57–64. Springer.

Dennis Assenmacher, Marco Niemann, Kilian Müller, Moritz Seiler, Dennis M Riehle, and Heike Trautmann. 2021. RP-Mod and RP-Crowd: Moderator-and crowd-annotated german news comment datasets. In *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Shiladitya Bhattacharya, Siddharth Singh, Ritesh Kumar, Akanksha Bansal, Akash Bhagat, Yogesh Dawer, Bornini Lahiri, and Atul Kr. Ojha. 2020. Developing a multilingual annotated corpus of misogyny and aggression. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 158–168, Marseille, France. European Language Resources Association (ELRA).

Wenzhi Cao, Vahid Mirjalili, and Sebastian Raschka. 2020. Rank consistent ordinal regression for neural networks with application to age estimation. *Pattern Recognition Letters*, 140:325–331.

Branden Chan, Stefan Schweter, and Timo Möller. 2020. German's next language model. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6788–6796, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Patricia Chiril, Véronique Moriceau, Farah Benamara, Alda Mari, Gloria Origgi, and Marlène Coulomb-Gully. 2020a. An annotated corpus for sexism detection in French tweets. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1397–1403, Marseille, France. European Language Resources Association.

Patricia Chiril, Véronique Moriceau, Farah Benamara, Alda Mari, Gloria Origgi, and Marlène Coulomb-Gully. 2020b. He said "who's gonna take care of your children when you are at ACL?": Reported sexist acts are not sexist. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4055–4066, Online. Association for Computational Linguistics.

Christoph Demus, Jonas Pitz, Mina Schütz, Nadine Probol, Melanie Siegel, and Dirk Labudde. 2022. A comprehensive dataset for german offensive language and conversation analysis. In *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)*, pages 143–153.

Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2018. Overview of the evalita 2018 task on automatic misogyny identification (ami). In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018)*.

Simona Frenda, Bilal Ghanem, Manuel Montes y Gómez, and Paolo Rosso. 2019. Online hate speech against women: Automatic identification of misogyny and sexism on twitter. *Journal of Intelligent & Fuzzy Systems*, 36(5):4743–4752.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. Datasheets for datasets. *Communications of the ACM*, 64(12):86–92.

Dylan Grosz and Patricia Conde-Cespedes. 2020. Automatic detection of sexist statements commonly used at the workplace. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Workshop (Learning Data Representation for Clustering) LDRC*.

Sarah Hewitt, T. Tiropanis, and C. Bokhove. 2016. The problem of identifying misogynist language on twitter (and other online social spaces). In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16, page 333–335, New York, NY, USA. Association for Computing Machinery.

Aiqi Jiang, Xiaohan Yang, Yang Liu, and Arkaitz Zubiaga. 2022. Swsr: A chinese dataset and lexicon for online sexism detection. *Online Social Networks and Media*, 27:100182.

Hannah Kirk, Abeba Birhane, Bertie Vidgen, and Leon Derczynski. 2022. Handling and presenting harmful text in nlp research. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 497–510.

Hannah Kirk, Wenjie Yin, Bertie Vidgen, and Paul Röttger. 2023. SemEval-2023 task 10: Explainable detection of online sexism. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2193–2210, Toronto, Canada. Association for Computational Linguistics.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229.

Hala Mulki and Bilal Ghanem. 2021. Let-mi: an arabic levantine twitter dataset for misogynistic language. *arXiv preprint arXiv:2103.10195*.

Pulkit Parikh, Harika Abburi, Pinkesh Badjatiya, Radhika Krishnan, Niyati Chhaya, Manish Gupta, and Vasudeva Varma. 2019. Multi-label categorization of accounts of sexism using a neural framework. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1642–1652, Hong Kong, China. Association for Computational Linguistics.

Johann Petrak and Brigitte Krenn. 2022. Misogyny classification of german newspaper forum comments. *arXiv preprint arXiv:2211.17163*.

Francisco Rodríguez-Sáchez, Jorge Carrillo de Albornoz, Laura Plaza, Julio Gonzalo, Paolo Rosso, Miriam Comet, and Trinidad Donoso. 2021. Overview of exist 2021: sexism identification in social networks. *Procesamiento del Lenguaje Natural*, 67.

Francisco Rodríguez-Sánchez, Jorge Carrillo-de Albornoz, and Laura Plaza. 2020. Automatic classification of sexism in social networks: An empirical study on twitter data. *IEEE Access*, 8:219563–219576.

Paul Röttger, Bertie Vidgen, Dirk Hovy, and Janet B Pierrehumbert. 2021. Two contrasting data annotation paradigms for subjective nlp tasks. *arXiv preprint arXiv:2112.07475*.

Niloofar Safi Samghabadi, Parth Patwa, Srinivas PYKL, Prerana Mukherjee, Amitava Das, and Thamar Solorio. 2020. Aggression and misogyny detection using BERT: A multi-task approach. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 126–131, Marseille, France. European Language Resources Association (ELRA).

Sima Sharifirad and Stan Matwin. 2019. When a tweet is actually sexist. A more comprehensive classification of different online harassment categories and the challenges in NLP. *CoRR*, abs/1902.10584.

Elena Shushkevich and John Cardiff. 2019. Automatic misogyniny detection in social media: A survey. *Computación y Sistemas*, 23(4).

Jane M Ussher. 2016. Misogyny. *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*, pages 1–3.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998—6008.

Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In *Proceedings of the NAACL student research workshop*, pages 88–93.

Philine Zeinert, Nanna Inie, and Leon Derczynski. 2021. Annotating online misogyny. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics.

A. Appendix: Sample comments

Content warning: In the following, examples of *sexist* comments are given as provided for training the annotators.

A sample comment is given for each criterion listed in Section 4.2 to judge a comment as *sexist*. The German examples are accompanied by their translation into English.

A.1. Generalizing stereotypes - attributions to groups of women.

A.1.1. Role stereotypes

Sample comment: ”Weiber sollen Kinder austragen und sich um diese sorgen! Pasta!Sonst sie zu nichts zu gebrauchen. Ein richtiges Weib im wahrsten Sinne de Wortes ist sowieso eine Göttin und der Mann göttlich. Feministinnen sind krank und unbrauchbar und sogar böse!”
EN translation: ”Women should carry children and take care of them! Pasta! Otherwise they are good for nothing. A real woman in the true sense of the word is a goddess anyway and the man divine. Feminists are sick and useless and even evil!”

A.1.2. Attribute stereotypes

Sample comment: ”Als ob körperliche Attraktivität bei Männern das maßgebliche Kriterium für eine sexuelle Beziehung wäre. Männer müssen reich und mächtig sein, Frauen schön. Das war seit jeher so.”
EN: ”As if physical attractiveness were the decisive criterion for a sexual relationship with men. Men have to be rich and powerful, women beautiful. It has been that way since the beginning of time.”

A.2. Reduction of a person to her appearance

Sample comment: ”Bei aller Tragik und Ernsthaftigkeit.... wir haben schon a fesche Justizministerin”
EN: ”With all tragedy and seriousness.... we definitely have a dashing Minister of Justice”

A.3. Women as sexual objects

Sample comment: ”Frauen wählten mehrheitlich Biden ... würden aber für Trump alle Hüllen fallen lassen.”
EN: ”Women voted in majority for Biden ... but would drop all their clothes for Trump.”

A.4. Female connoted insult

Sample comment: ”Ich mag keine Kampflesben,die sollte man mal allesamt wegsperren”
EN: ”I do not like combat lesbians, they should all be locked away”

A.5. Denigration of women, their performance and women’s issues

Sample comment: ”Frauen benutzen, Autos, Smartphones, fahren über Brücken, die nicht einstürzen. Aber Frauen machen keine Autos, sie machen keine Smartphones, kaum eine Frau versteht, warum eine Brücke stehen bleibt und nicht einstürzt.”
EN: ”Women use, cars, smartphones, drive over bridges that do not collapse. But women do not make cars, they do not make smartphones, hardly any women understand why a bridge stands and does not collapse.”

A.6. Downplay sexual violence and sexual harassment against women

Sample comment: ”Jetzt wissen wir wenigsten welche Filmsternchen sich durchs Bett hochgearbeitet haben. Nach 25 Jahren stockt die Karriere also wird jetzt verklagt. Diese miese #Metoo Hinrichtungskamapgne wird kläglich untergehen. Aber medial ist sie ein toller Erfolg.”
EN: ”Now we know at least which movie starlets have worked their way up through the bed. After 25 years the career stalls so now they sue. This lousy #Metoo execution camapgne will go down miserably. But media-wise it is a great success.”

A.7. Whataboutism

Sample comment: ”Lob für Frauenarbeit? Männer bringen auch einen sehr großen Einsatz und das wird an den Opferzahlen deutlich: Die meisten Opfer sind Männer!!! Warum muss diese Unterscheidung immer sein? Ist das Selbstbewusstsein von Frauen so gering ausgebildet?”
EN: ”Praise for women’s work? Men also do a great deal of work and that is clear from the numbers of victims: most of the victims are men!!! Why does this distinction always have to be? Is women’s self-awareness so poorly developed?”

A.8. Abortion

Sample comment: ”Also zuerst einmal ist Abtreibung gleich Mord. Das heißt Abtreibung sollte genauso bestraft werden wie Mord. Selbstbestimmung hat seine Grenzen. Es darf nicht ein anderes Lebewesen gefährdet werden egal wie selbstbestimmt man leben möchte.”
EN: ”So first of all, abortion equals murder. That means abortion should be punished the same as murder. Self-determination has its limits. Another living being must not be endangered no matter how self-determined one wants to live.”

A.9. Misandry

Sample comment: ”Wie der zweite Weltkrieg zeigte, sind tote Männer das Beste, was einer Gesellschaft passieren kann.”
EN: ”As World War II showed, dead men are the best what can happen to a society.”