Gender-tailored Semantic Role Profiling for German

Manfred Klenner, Anne Göhring, Alison Yong-Ju Kim, Dylan Massey
Department of Computational Linguistics
University of Zurich
{klenner,goehring}@cl.uzh.ch

Abstract

In this short paper, we combine the semantic perspective of particular verbs as casting a positive or negative relationship between their role fillers with a pragmatic examination of how the distribution of particular vulnerable role filler *subtypes* (children, migrants, etc.) looks like. We focus on the *gender* subtype and strive to extract gender-specific semantic role profiles: who are the predominant sources and targets of which polar events - men or women[1]? Such profiles might reveal gender stereotypes or biases (of the media), but as well could be indicative of our social reality.

1 Introduction

Some verbs express a positive or negative relationship (a polar relation) between the fillers of their semantic roles. For example, we can infer from the sentence "He offended her," an even, reciprocally holding, negative relation (e.g. *against(he,her)*. Moreover, such semantic roles might bear a polar (i.e. positive or negative) load, e.g. the agent of cheating might be regarded as a negative actor, a villain. From a pragmatic point of view, it might be interesting to take a closer look at the distribution of particular role fillers or role filler groups of such verbs indicating a polar relation, namely vulnerable groups such as children (pedophilia), migrants (xenophobia), people of color (racist bias), and certain gender identities (gender bias). This could reveal interesting facts about the conceptualization and contextualization of these filler groups in the real world. Such an approach could be useful for various kinds of monitoring processes (e.g. discrimination motoring). In this short paper, we focus on gender. Our goal is to enable gender-tailored semantic profiling. On a micro level, semantic profiling strives to identify the roles that gender denoting nouns occupy, e.g. that female nouns occur quite often as patients (targets) of physical violence, while male denoting nouns often are filler of the patient role of torture or accusation. On the macro level, a general, cross-verb inventory of semantic roles like *villain, victim, benefactor, beneficiary* could be used to aggregate gender-specific conceptualization. Here, we focus on the micro level.

We introduce a classifier that determines the grammatical gender of human-denoting German nouns. We combine this with our rule-based sentiment inference system[2] (Klenner et al., 2017) which assigns two types of relations between entities: in favor of, against. Each verb of our verb lexicon expresses such a polar relation and has a source (the agent) and a target (the patient, recipient or theme) role. We filtered the output of our system for cases in which the gender classifier labeled at least one of the verb roles as male- or female-denoting[3]. With such data, we were able to filter for polar events in which men are sources and women targets (and vice versa). On the basis of statistical tests, cases are found in which female or male denoting nouns are significantly over- or underrepresented.

2 Related Work

Currently, gender classification is primarily restricted to predicting the gender of text authors of blogs, see Mukherjee and Liu (2010), or to find out whether a headline is about a man or a woman, see Campa et al. (2019).

Sun and Peng (2021) observe a gender-specific tendency to combine personal and professional events in the Wikipedia pages of celebrities, an

[1] Certainly, we do not claim that gender is a binary category; but gender-denoting nouns without explicit indications (e.g. '*') do have a binary reference that we cannot overcome.

[2] The online version can be found here: https://pub.cl.uzh.ch/demo/stancer/index.py.

[3] Thus, there is no need to assign semantic roles explicitely.

asymmetric association where e.g. women's personal events appear more often in the career section than for men. They also establish higher efficiency when extracting events (verb denotations) over analyzing raw text for detecting this gender bias.

Bias detection and debiasing, in general, are important research topics (see Stanczak and Augenstein (2021) for a survey). Researchers use metrics such as pointwise mutual information (PMI) to measure the association of words with gender (Stanczak et al., 2021). We look into cases in which both grammatical genders co-occur with a verb, i.e. when PMI cannot be used.

3 Grammatical Gender Classification

The basis for our gender classifier is the freely available resource (Klenner and Göhring, 2022) of 13,000 German nouns which were manually classified as denoting either animate or inanimate entities[4]. In order to create a gold standard for grammatical gender classification, we took a subset containing animate singular nouns and manually[5] selected those that can be used to refer to women or men (altogether 4,320). Examples of female-denoting nouns include *Schwester, Gastgeberin, Schauspielerin* (Eng. *sister, hostess, actress*, respectively). We then saw that the data was imbalanced, namely that there were more male-denoting nouns (2,830) than female-denoting ones (1,490). As such a dataset would have produced a biased classifier with better classification for male-denoting nouns, we searched for more female-denoting nouns, ultimately expanding this set to 3,700. In German, this can generally be carried out by adding the suffix *in* to the end of male-denoting nouns, e.g. *Helfer* → *Helfer**in*** (Eng. *helper*). If such a variant is found in a corpus, it is added to the female list. Since we found that female nouns in news texts are underrepresented, we decided to keep the whole list of 3,700 female nouns for learning.

In Klenner and Göhring (2022) we tested various word embeddings (GloVe, BERT,FastText) for the training of the animacy classifier (MLP, SVM, LR) and found FastText with logistic regression (LR) to perform best. Therefore, we used only FastText embeddings to train a LR model for gender-aware animacy classification. There was no need to carry out extensive experiments, since our initial model achieved a high accuracy of 97.29%. Table 1 shows the results of a random 75/25 train/test split. Female-denoting noun identification with a precision of 98.3% and a recall of 99.3% might help us to mitigate gender imbalance in news texts.

	non-actors	female	male
precision	0.967	0.983	0.973
recall	0.984	0.993	0.927
f1	0.975	0.988	0.949

Table 1: Performance of our three-way, gender-aware animacy classification model.

Note: Not all German female-denoting nouns possess the "in" ending. In fact, in our list of female-denoting nouns, 50 have endings other than "in" (e.g. *Frisöse*, Eng. hairdresser). A rather simple (rule-based) method was to classify a word with an "in" ending as a female-denoting noun. But that would produce quite some error. In a corpus of 25 million nouns, we found 67,823 words (tokens) ending with "in". For 36,247 cases of these "in"-words our classifier predicted *female*. The remaining 31,576 "in"-nouns correspond to 4,035 types. We manually classified 1,000 and found only 5 female-denoting words. Thus, the classifier does not base its decision on the suffix, though this would be a legitimate approach since FastText uses sub-word splitting. The performance of our classifier with respect to the non-"in" female-denoting nouns cannot reliably be evaluated at the moment. We leave it to future work to train models able to deal with these rarer cases.

4 Statistical Setting

Our question of interest was that of identifying an imbalance, if any, between men and women, or some gender-specific behavioral semantic profile, as portrayed in newspaper texts. We focused on men and women's roles as positive or negative actors (sources) or as being positively or negatively affected patients (targets). In particular, we looked at all polar verb instantiations, with male- and female-denoting nouns occupying the source and target roles. Then, we gathered statistics on how often a positive or negative relation between two gender-denoting nouns (e.g. a female- and a male-denoting noun) was found. We performed this for all gender permutations at the level of a single verb, but we also accumulated this over all verbs. To evaluate whether a verb is more biased

[4]download: https://zenodo.org/record/7630043#.Y-aCU9LMJH4

[5]The annotation task is straightforward for a native speaker; thus, only one annotator was needed.

towards male or female roles, the (prior) gender distribution in the whole data must be taken into account. In our text corpus, we found a ratio of male- (1,290,415) to female- (283,952) denoting nouns (according to the gender classifier) of about 4:1. That is, the maximum likelihood estimated probability of male denoting nouns is 0.815, that of female 0.185.

The data is binomially distributed for each role of a verb frame. For instance, if a transitive (active voice) verb has $n = 200$ instantiations (and thus 200 sources), of which 20 are female, then we determine the cumulative probability of up to 20 cases given 200 trials with $p = 0.185$ as $\sum_{i=1}^{20} binom(i, 200, 0.185)$. If this value is below $\alpha = 0.05$, then we reject H_0 and adopt H_1, i.e. we can conclude that the verb (usage) is biased, and similarly for the $1 - 0.95\%$ interval. Spelled out, H_0 claims that female (male) denoting nouns occupy source (target) verb roles according to their prior probability. If this is for some verbs rather unlikely, than H_1 is adopted saying there is a verb-role specific bias, for instance that female denoting nouns are significantly more often targets of (verbs of) physical violence than male denoting nouns.

We only looked into verbs for which a normal distribution could be approximately assumed, which is given if $n * p \geq 5$ and $n(1 - p) \geq 5$, where n is the number of cases. In our setting, this amounts to a frequency threshold of $n = 5/0.185 = 27$. For each verb above this frequency threshold, we tested the null hypothesis H_0 that male- and female-denoting nouns occupy the role of a verb according to their respective distributions in the whole corpus.

5 Empirical Results

We use data from 3 Swiss newspapers published between 2004 and 2014. Despite the medium corpus size, the cases in which a verb has 2 animate role fillers (singular male or female[6]) at the source and target positions of that verb are relatively infrequent. This low frequency can be attributed to (1) the abundance of cases written in passive voice (for which there is quite often no source indicating PP) and (2) cases in which the source or holder is a personal pronoun (which, in German, leaves the animacy status of the referent open). In German, inanimate objects might have non-neutral grammatical gender, e.g. German *Brücke* (Eng. *bridge*) is feminine. This reduces the number of instantiations, e.g. for the verb *töten* (Eng. *to kill*) the counts shrink from 26,200 to 1,110 (21,000 passive cases, 4,100 pronouns). As gender classification is done after sentiment inference, another 800 cases disappear since no or only one gender-denoting noun was found, ultimately leaving 302 cases of *töten*.

[6]We did not take plural nouns into account since German plural male nouns for a long time have been regarded as being generic, denoting all genders. The gender reference of such a noun, thus, cannot be reliably fixed.

relation	source	target	#
+	male	male	30
+	male	female	5
+	female	male	6
+	female	female	2
-	male	male	1273
-	male	female	**707**
-	female	male	**221**
-	female	female	63

Table 2: Overview: number of positive (+) and negative (-) relations between the gender referring nouns.

5.1 1st Experiment: Source Imbalance

From the output of the sentiment inference system for these texts, 132 verb types display cases of an animate source *and* target. Only 20 verbs pass the strict threshold (≥ 27), and of these, 10 have a gender-specific imbalance. Table 2 shows the overall statistics. We can see that negative relations from a male-denoting noun (as source) to a female-denoting noun (as target) occur about 3 times as often as the other way around (in bold).

If we observe the most frequent verbs of these two bidirectional cases, it turns out that they are gender-specific. Among verbs whose sources are female-denoting nouns, the most frequent are (in ascending order) *coerce, deceive, threaten, accuse*; for male-denoting noun sources: *attack, kill, rape*.

Table 3 shows the list of 10 verbs with gender-specific source-role imbalance. For 7 of these verbs, male-denoting nouns take on the source role significantly more often than expected (the error risk α is 5%). A letter f (m) in a column $\leq$ means that the probability of #f (#m) female (male) sources for the verb is less than or equal to α.

In order to quantify the noise in our empirical analysis, we manually inspected all cases from Table 3. We looked for gender classification and sentiment relation errors. The last column (#e) in Table

verb	$\leq$	$\geq$	#f+m	#f	#m	#e
attack	f	m	62	5	57	7
harass	m	f	76	31	45	1
fire	f	m	157	17	140	5
shot dead	f	m	194	25	169	1
critizice	f	m	33	1	32	3
kill (töten)	f	m	302	23	279	20
kill	f	-	62	6	56	1
rape	f	m	46	2	44	3
indict	-	f	30	9	21	0
assault	f	m	60	5	55	6

Table 3: Gender specific source role imbalance (f=female, m=male, $\leq$ means $\leq \alpha$, $\geq$ means $\geq 1-\alpha$, e=prediction error

3 shows the error counts. E.g. *attack* was associated with 5 cases of animals in the source role and 2 cases of generic male plural nouns, which can also be used as a feminine singular noun (*Unbekannte*, Eng. unknown females). A manual analysis revealed an error rate of 4.6% (47 out of 1022).

5.2 2nd Experiment: Target Imbalance

As stated, the low frequencies shown in Table 2 are partly due to the high number of passive sentences, in which typically only a target can be found. However, we can also perform statistical tests with targets only, which would help us determine whether men and women are significantly less or more often targets than their respective distributions suggest. We found 793,246 instantiations of 233 verbs in passive voice, 66 for which we found gender-specific patterns. For instance, men are more often target of torture (line *foltern* in the appendix), *verwunden* (Eng. *injury*), *verdächtigen* (Eng. *suspect*), and *anklagen* (Eng. *accuse*) than women, who more often are targets of *vergewaltigen* (Eng. *rape*), *zwingen* (Eng. *coerce*), *benachteiligen* (Eng. *discriminate*), and *erniedrigen* (Eng. *humiliate*).

5.3 3rd Experiment: Inanimate Targets

We also tried to identify the inanimate objects (targets) toward which men and women hold a favorable or opposing view (e.g. *lies* in *She detests lies*). At the token level, we have: 3,180 +f, 1,477 -f, 22,689 +m and 9,935 -m (e.g. 9,935 negative attitudes of male towards something). At the type level: 1,857 +f, 1,030 -f, 7,258 +m, 4,564 -m. Still, the ratio of men:women is imbalanced: there are far more male- than female-denoting sources. Table 4 shows the word-level intersection percentage of the target topics that we found. The intersection is not high. A close inspection might reveal interesting differences; we leave this for future work.

	f	m	$\cap$	%
+	1857	7258	944	10.3
-	1030	4564	500	8.9

Table 4: Men and women: likes (+) and dislikes (-) ($\cap$ =intersection)

5.4 4th Experiment: Polar Targets

One final experiment again deals with inanimate targets, but this time we look how often men and women are *in favor* of something positive or negative, and correspondingly for the *against* relation. For this task, we use our polarity lexicon Clematide and Klenner (2010)[7], albeit without NP sentiment composition; only words are used. Table 5 shows the results. For instance, there are 1,242 cases in which men are against something positive ($-\ \rightarrow$pos), e.g. decriminalization or democracy. In this paper, we have discussed the methods to generate these candidates, future work is devoted to a fine-grained qualitative analysis.

gender	+ $\rightarrow$pos	+ $\rightarrow$neg	- $\rightarrow$pos	- $\rightarrow$neg
female	149	144	178	35
male	944	896	1242	214

Table 5: In favour of + and -, against + and -, gender-specific, where pos/neg is a positive/negative word

6 Conclusion and Outlook

We introduced gender-tailored semantic role profiling on the basis of grammatical gender detection and sentiment relation extraction. Our model combines the first classifier for the detection of the grammatical gender of German nouns with an existing rule-based sentiment relation extractor. In a case study, we were able to carve out the different semantic role profiles of male and female denoting expressions in news texts from 2004 to 2014. In more recent work, we have compared the analysis of the data from 2004 to 2014 to results from the same newspapers from 2015 to present-day (Klenner, 2023), in order to see whether semantic profiles have changed or not.

[7]See under: "PolArt"-Lexicon from https://sites.google.com/site/iggsahome/downloads.

7 Discussion of Limitations

Our method detects gender imbalance by using an existing rule-based system and a new grammatical gender classifier. Neither performs perfectly, and we cannot claim that our sampling methods produce representative data drawn from the whole population. Rather, we work with a subset that can be identified by our tools. Generalizing from the subset to the population is not our intention; rather, our approach is a first step toward gender-tailored sentiment analysis. Finally, we do not claim to find biases in the data, but instead speak of imbalance and propose that a qualitative analysis of the results is needed.

8 Appendix: Table of Target Imbalance

verb	#	#f	#m	tendency
anklagen	753	89	664	$\geq$ m $\leq$ f
anzeigen	324	42	282	$\geq$ m $\leq$ f
bedrängen	96	33	63	$\leq$ m $\geq$ f
belästigen	83	38	45	$\leq$ m $\geq$ f
benachteiligen	65	21	44	$\leq$ m $\geq$ f
beschuldigen	492	60	432	$\geq$ m $\leq$ f
beschützen	33	10	23	$\geq$ f
bestehlen	49	17	32	$\leq$ m $\geq$ f
bestrafen	1093	153	940	$\geq$ m $\leq$ f
betrügen	136	40	96	$\leq$ m $\geq$ f
demütigen	57	16	41	$\leq$ m $\geq$ f
deportieren	76	24	52	$\leq$ m $\geq$ f
diffamieren	29	2	27	$\geq$ m
diskriminieren	65	24	41	$\leq$ m $\geq$ f
drohen	611	136	475	$\leq$ m $\geq$ f
einschüchtern	28	9	19	$\leq$ m $\geq$ f
foltern	224	30	194	$\geq$ m $\leq$ f
kritisieren	378	45	333	$\geq$ m $\leq$ f
misshandeln	108	31	77	$\leq$ m $\geq$ f
nötigen	84	25	59	$\leq$ m $\geq$ f
töten	2385	383	2002	$\geq$ m $\leq$ f
umbringen	329	71	258	$\leq$ m $\geq$ f
unterdrücken	41	17	24	$\leq$ m $\geq$ f
verdächtigen	580	70	510	$\geq$ m $\leq$ f
vergewaltigen	286	213	73	$\leq$ m $\geq$ f
verwunden	100	10	90	$\geq$ m $\leq$ f
vorwerfen	1154	142	1012	$\geq$ m $\leq$ f
widerlegen	55	17	38	$\leq$ m $\geq$ f
zwingen	985	232	753	$\leq$ m $\geq$ f
überfallen	174	52	122	$\leq$ m $\geq$ f

Table 6: Statistically significant target role imbalance: $\leq m \geq f$ means: men are significantly fewer target (then their distribution in the data suggests), women significantly more often. Other case accordingly.

Acknowledgements

Our work was supported by the Swiss National Foundation (SNF) under the project number 105215_179302 from 2018 to 2022.

References

Stephanie Campa, Maggie Davis, and Daniela Gonzalez. 2019. Deep and machine learning approaches to analyzing gender representations in journalism. Online.

Simon Clematide and Manfred Klenner. 2010. Evaluation and extension of a polarity lexicon for German. In *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 7–13.

Manfred Klenner. 2023. Sentiment inference for gender profiling. In *Proceedings of the 4th Conference on Language, Data and Knowledge*. in press.

Manfred Klenner and Anne Göhring. 2022. Animacy denoting german nouns: Annotation and classification. In *Proceedings of the Language Resources and Evaluation Conference*, pages 1360–1364, Marseille, France. European Language Resources Association (ELRA).

Manfred Klenner, Don Tuggener, and Simon Clematide. 2017. Stance detection in Facebook posts of a German right-wing party. In *LSDSem 2017/LSD-Sem Linking Models of Lexical, Sentential and Discourse-level Semantics*.

Arjun Mukherjee and Bing Liu. 2010. Improving gender classification of blog authors. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 207–217, Cambridge, MA. Association for Computational Linguistics.

Karolina Stanczak and Isabelle Augenstein. 2021. A survey on gender bias in natural language processing. *CoRR*, abs/2112.14168.

Karolina Stanczak, Sagnik Ray Choudhury, Tiago Pimentel, Ryan Cotterell, and Isabelle Augenstein. 2021. Quantifying gender bias towards politicians in cross-lingual language models. *CoRR*, abs/2104.07505.

Jiao Sun and Nanyun Peng. 2021. Men are elected, women are married: Events gender bias on Wikipedia. In *Proceedings of the 59th Annual Meeting of the ACL and the 11th International Joint Conference on Natural Language Processing*, pages 350–360, Online. Association for Computational Linguistics.