A Checkpoint on Multilingual Misogyny Identification

Arianna Muti and **Alberto Barrón-Cedeño**
Department of Interpreting and Translation
Alma Mater Studiorum–Universitá di Bologna
Forlì, Italy
{arianna.muti2,a.barron}@unibo.it

Abstract

We address the problem of identifying misogyny in tweets in mono and multilingual settings in three languages: English, Italian and Spanish. We explore model variations considering single and multiple languages both in the pre-training of the transformer and in the training of the downstream task to explore the feasibility of detecting misogyny through a transfer learning approach across multiple languages. That is, we train monolingual transformers with monolingual data and multilingual transformers with both monolingual and multilingual data. Our models reach state-of-the-art performance on all three languages. The single-language BERT models perform the best, closely followed by different configurations of multilingual BERT models. The performance drops in zero-shot classification across languages. Our error analysis shows that multilingual and monolingual models tend to make the same mistakes.

Disclaimer: Due to the nature of the topic, this paper contains offensive words.

1 Introduction

Misogynous contents express hate towards women in the form of insulting, sexual harassment, male privilege, patriarchy, gender discrimination, belittling, violence, body shaming and sexual objectification (Srivastava et al., 2017). According to a study by *Vox–Osservatorio Italiano sui diritti* on hate speech against minorities (women, homosexuals, migrants, people with disabilities, Jews and Muslims) in Italian tweets,[1] women are the most targeted group. They observed a significant increase in the number of misogynous tweets from 2019 to 2021: shifting from 26% to 44% of all hateful posts. Blake et al. (2021) observed a correlation between misogyny on Twitter and domestic violence in specific areas has, stressing the importance of flagging such contents to try to dim their impact online.

We target the problem of identifying misogyny in multiple languages. This work represents a first step towards investigating the specificity of misogyny with respect to language and culture. To address this novel research question, we test two hypotheses:

H1 More data boosts the model performance, even if it is in a different language; therefore, considering training material in diverse languages benefits in the prediction of misogyny in such languages.

H2 misogyny is language-specific and therefore a monolingual model performs better, even if it is trained on smaller data.

We rely on: (a) data in each of the languages in isolation; or (b) data in various languages in conjunction, through the training of a single multilingual model.[2] We exploit monolingual transformers (BERT (Devlin et al., 2019)) for three languages — English, Italian, and Spanish — and one multilingual transformer (Multilingual BERT (Devlin et al., 2019)). We perform a thorough exploration combining different settings, which include training monolingual transformers with monolingual data, multilingual transformers with monolingual data, and multilingual transformers with multilingual data.

Section 2 summarizes the related work on misogyny identification, both in mono- and multilingual settings. Section 3 describes the datasets. Section 4 describes the methodology, whereas Section 5 discusses the obtained results. Section 6 shows our

[1] http://www.voxdiritti.it/la-nuova-mappa-dellintolleranza-6/

[2] Our settings avoid resorting to machine translation because the jargon used to convey hateful messages tends to produce faulty target texts, causing the classifiers to struggle (Casula and Tonelli, 2020; Pamungkas and Patti, 2019).

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Student Research Workshop, pages 454 - 460

error analysis. Section 7 and 8 provide a conclusion and an overview on the societal impact and limitations of our work.

2 Related Work

Monolingual Approaches The increasing number of hateful posts against women has attracted the interest of the scientific community, but it remains an underexplored field compared to other types of hate speech (Tontodimamma et al., 2021). Work on automatic misogyny identification has been carried out in a limited number of languages. For instance, the Automatic Misogyny Identification (AMI) series of shared tasks launched in EVALITA (Fersini et al., 2018, 2020) and IberEval (Anzovino et al., 2018) has produced evaluation frameworks to identify misogynous tweets in English, Italian and Spanish. HatEval at SemEval 2019 (Basile et al., 2019) focused on the detection of hate speech towards women and immigrants in English and Spanish.

Participants in those shared tasks mostly used TF–IDF representations (e.g., Frenda et al. (2018)), word embeddings (e.g., Fabrizi (2020)), and sentence embeddings (e.g., Ahluwalia et al. (2018)). When extracting lexical features from social media, it is common to represent hashtags, emoticons and mentions as well. For instance, Pamungkas and Patti (2019) considered both a bag of hashtags and a bag of emojis. They also encoded information about the occurrence of swear words.

Among the most commonly-used classifiers there are recurrent neural networks (Goenaga et al., 2018; Buscaldi, 2018), convolutional neural networks (da Silva and Roman, 2020), shallow models (Pamungkas et al., 2018) and transformer-based models (Lees et al., 2020; Muti and Barrón-Cedeño, 2020), which perform the best.

Multilingual Approaches Few works are focused on the multilingual identification of misogyny. Basile and Rubagotti (2018) adopted a bleaching approach, i.e. transforming lexical strings into more abstract features (van der Goot et al., 2018), and tested their model on Italian and English. They use an SVM with n-gram features. This is a close work to ours: they train on L1 and test on L1, train on L2 and test on L2, and they also train and test on both languages in combination.

Pamungkas and Patti (2019) created bilingual misogynist data in English, Italian and Spanish with machine translation to train in a source language and predict in a target language with an LSTM. We neglect the use of machine translation at all stages.

	training		testing	
Dataset	**not mis**	**mis**	**not mis**	**mis**
en EVALITA 2018	2,215	1,785	540	460
es IberEval 2018	1,658	1,649	416	415
it EVALITA 2018	2,171	1,828	509	491

Table 1: Class distribution for the three corpora in English (en), Spanish (es) and Italian (it).

Pamungkas et al. (2020) adopted an approach similar to ours, using multilingual transformers to identify English, Spanish and Italian misogynist tweets. The difference is that they only train a model on one language and test it on a different one, without considering all language combinations.

Differently from the previous works, we do not focus on model performance or engineering, but we head toward investigating a novel research question: is misogyny language-specific?

3 The multi-AMI Evaluation Framework

We consider misogyny datasets in three languages, released under two editions of the AMI shared task: AMI at IberEval 2018 (Anzovino et al., 2018) and AMI at EVALITA 2018 (Fersini et al., 2018). AMI at IberEval 2018 focused on identifying misogyny on English and Spanish tweets, and in classifying misogynistic instances in different categories. AMI at EVALITA 2018 focused on two tasks in Italian. Task A addressed misogyny identification, whereas Task B aimed at recognizing whether a misogynous tweet is person-specific or generally addressed towards a group of women. We address the binary problem alone: whether a tweet is misogynist or not. Table 1 shows statistics for the three corpora.

We stick to the evaluation metric of AMI: the F_1 measure. For direct comparison with our models we consider the best-performing approaches in both shared tasks. For Italian, Bakarov (2018) used TF-IDF weighting combined with singular value decomposition and an ensemble of classifiers. For English, Saha et al. (2018) concatenated sentence and average word embeddings with TF–IDF weights coupled with a logistic regression model. For Spanish, Pamungkas et al. (2018) applied an SVM with a series of lexical features, including lexicons of abusive words. The bottom row of Table 2 shows the performance of the three models.

4 Model Description

Our models to identify misogynous tweets are built on different variations of BERT (Devlin et al., 2019). In the monolingual settings, we use bert-base-uncased for English (Devlin et al., 2019), bert-base-spanish-wwm-uncased for Spanish (Cañete et al., 2020) and AlBERTo for Italian (Polignano et al., 2019). For the multilingual settings, we use multilingual BERT (mBERT) (Devlin et al., 2019). mBERT has the same architecture as BERT, but it is trained on Wikipedia articles in multiple languages (Liu et al., 2020). We also apply mBERT in monolingual settings, to observe its behaviour in zero-shot classification across languages.

Our output layer is a soft-max with two units. We use the categorical cross-entropy loss function and the AdamW optimizer with a learning rate of 1-8 (Loshchilov and Hutter, 2017), batch size of 16 and 4 training epochs.

train	**en**	**es**	**it**	**all**
BERT en	**0.71**	–	–	–
BERT es	–	**0.85**	–	–
BERT it	–	–	**0.87**	–
mBERT en	0.65	0.14	0.17	–
mBERT es	0.62	0.81	0.50	–
mBERT it	0.47	0.63	0.87	–
mBERT en-es	0.67	0.83	–	0.75
mBERT en–it	0.66	–	0.86	0.77
mBERT es–it	–	0.80	0.86	0.84
mBERT en–es–it	0.68	0.82	0.86	0.78
best-AMI	0.70	0.81	0.84	–

Table 2: F_1 performance for the different language combinations. Best AMI shared task models shown at the bottom for comparison (cf. Section 3).

5 Experiments and Results

Our objective is to assess whether and to what extent considering training material in diverse languages benefits in the prediction of misogyny in multiple languages. We carried out a number of experiments to test hypotheses H1 and H2 (cf. Section 1).

We head toward investigating the way in which misogyny is expressed in different languages. Even if the impact of shared vocabulary in multilingual settings remains unclear (Liu et al., 2020), we explore the feasibility of using multilingual embeddings to produce zero-shot classifications across languages —training on L_1 to predict on L_2— and as a data augmentation technique —training on L_1+L_2 to predict on L_1.

We trained ten models considering all combinations of data in English (en), Spanish (es) and Italian (it): (*i*) one BERT model per language, (*ii*) one mBERT model per language, (*iii*) one mBERT model per each language pair, and (*iv*) one mBERT model with all three languages. Table 2 shows the results when predicting on data in each language and all together. The scores under columns en, es and it are comparable, whereas those under all are not, because the testing sets are different.

The monolingual BERT models consistently perform the best, improving over the best AMI approaches (cf. Section 3). There is a performance drop when monolingual models are trained on top of mBERT, with the model trained on English achieving the poorest performance: as low as F_1=0.14 and 0.17 when tested on Spanish and Italian and six points lower on English than the monolingual BERT alternative. The results suggest that this transfer learning approach is not suitable for languages which are relatively far from each other, e.g., a Romance and a Germanic one. Considering a second language during training improves the predictions of the mBERT models (*i*) on English in all three cases, (*ii*) on Spanish with pair en–es, but (*iii*) not on Italian. Indeed, combining English and Spanish produces better results for both languages than when combining either with Italian. Considering all three languages results in mixed effects. It has the best mBERT performance on English, but runs short by one point with respect to the pairwise combinations on the other two languages. The best performance on all three languages together is obtained when neglecting the training data in English: F_1=0.84.

These results confirm H1 only partially. On the one hand, monolingual models built on top of a monolingual BERT performs the best. On the other hand, considering multilingual training data with a multilingual BERT improves over considering monolingual data alone.

We performed an additional experiment to verify that the performance shifts are not caused by the increase in the volume of training data, rather than the inclusion of another language. We trained a bilingual English–Italian model considering only 2,000 instances per language (conforming to the volume of the monolingual datasets). The performance on the English test set drops from F_1=0.65

it	FN	FP	en	FN	FP	es	FN	FP
bel	1	17	hysterical	27	20	puta	16	25
tette	0	8	woman	16	33	polla	3	13
culo	0	12	women	12	35	cállate	0	6
culona	3	12	fuck	9	27	acoso	7	5
porca	6	0	pussy	5	23	callate	2	4
figa	0	8	rape	3	27	madre	3	3
cazzo	3	4	bitch	4	22	mujer	6	5

Table 3: The most common words (sorted by inverse frequency) with the number of false positives and negatives in which they occur in the monolingual settings.

it	FN	FP	en	FN	FP	es	FN	FP
culo	2	16	hysterical	28	20	puta	24	25
bel	2	20	woman	19	34	polla	2	25
figa	2	11	women	28	31	cállate	6	5
cazzo	0	7	fuck	3	37	callate	1	8
troia	7	3	rape	4	39	madre	4	3
tette	3	3	fucking	4	29	acoso	5	9
culona	1	12	bitch	3	26	escoria	1	7

Table 4: The most common words (sorted by inverse frequency) with the number of false positives and negatives in which they occur in the multilingual settings.

to 0.54; on Italian it does from 0.87 to 0.85.

These results play in favour of H2: with the same amount of training data, the models do not benefit from data in other languages. Although this hints that H2 is true, these experiments are not enough to prove that misogyny is language-specific. The results obtained with the mBERT models when trained on single languages — no difference when compared against BERT models on Italian, but a drop of six and four points on English and Spanish— give more confidence that H2 might be true.

6 Error Analysis

We conducted an error analysis to assess how and which kind of errors are transferred from the mono- to the multilingual setting. We want to answer two questions. Question Q1 allows observing the behavior of the multilingual model with respect to the monolingual ones. Question Q2 helps to identify the words that are most likely responsible for the misclassification in the three languages.

Q1 Which instances are classified differently by the monolingual and the multilingual model?

The number of false positives and false negatives behave similarly in all languages. We discuss instances in English for the sake of clarity. We analyse the instances that the monolingual model (BERT en) classified correctly and the multilingual one (mBERT en-es-it) got wrong. We find 122 instances, with 51 false negative (FN) and 71 false positive (FP). Among the FN, the five most common lexical words are *hysterical*, *woman*, *skank*, *women* and *ass*. Among the FP, the words *rape* and *women* are very present, followed by *fucking*, *fuck* and *shut*. We notice that FN instances are more lexically diverse.

We also observe the intersection of misogynist tweets between the two models. The mono and multilingual model judged 543 and 541 tweets as misogynist. The intersection is of 438 instances, with 307 being correctly identified. Therefore, the majority of misogynist instances are detected by both models. This hints that there is no big difference between the models.

es	en	it	en
acoso	harassment	bel	beautiful
callate	shut up	cazzo	dick
cállate	shut up	culo	ass
escoria	scum	culona	big ass
madre	mother	figa	pussy
mujer	woman	porca*	slut
polla	dick	tette	boobs
puta	whore	troia	whore

*in most of the cases it refers to the expression *porca puttana* ≈ holy shit.

Table 5: Translation of the most common words in both Spanish and Italian into English.

Q2 Which words are most present in instances misclassified by both mono and multilingual models?

We first observe instances misclassified by both models. We find 70 FNs and 131 FPs. Table 3 shows the most frequent words in misclassified tweets in the mono- and Table 4 in the multilingual settings. Table 5 shows the translations of the Spanish and Italian words. No significant differences are observed across datasets of the same language, but there are big differences in how misogyny is expressed. In Italian, most words are related to the physical appearance of a woman, linked to sexual objectification. Italian language shows more linguistic creativity. Whereas English contains more insults, Spanish is more aggressive.

For English, the most frequent words are the

1	La ragazza che lavora nel negozio dove vado a fare sempre shopping mi ha detto che ho un bel culo :3333 *The girl working at the place where I always do shopping told me I have a nice ass.*
2	He said she said di Ashley Tisdale fa uscire il puttanone che è in me *He said she said by Ashley Tisdale brings out the bitch in me.*
3	ciao kikka buon pm quanto 6 figa e sexy [. . .] *hi kikka good evening you are so hot and sexy*
4	figa stai zitta che sono a casa da sola *oh don't tell me I'm home alone*

Table 6: Instances of tweets misclassified by both the monolingual and multilingual models (original Italian tweets followed by English translations).

same in both settings: *hysterical*, *woman*, *women*, *fuck* and *rape*. The fact that *woman* and *women* lie in the second and third positions might indicate an unintended identity-term bias (Fersini et al., 2020), for which the model learnt that *woman* occurs in misogynistic contexts. In both cases, the words *hysterical*, *rape* and *kitchen* (linked to women's stereotyped role) have a strong co-occurrence with the terms *women, woman*, therefore we can assume that these words trigger an error. The word *rape* is common in highly offensive contexts, making it a decisive feature for misogyny; it is frequently present in false positives.

For Spanish, words *puta*, *polla* and *cállate* are common for both settings. We focus the rest of the analysis on Italian, since it shows the biggest discrepancies. Table 6 shows examples. In both cases, *bel* always co-occur with *culo*. In FPs, it is commonly used by women to comment on themselves in a positive way, as in example 1. The same happens with the word *tette*, where in FP instances women usually complain about their breast size. These words tend to occur in offensive contexts and therefore are inclined to be classified as misogynist. Another interesting phenomenon that triggers FPs is the presence of slur reappropriation, i.e. women reclaiming certain negative terms (Felmlee et al., 2020), as in example 2 of Table 6. Another word that triggers FPs is *figa*, as it is typically used in hypersexualised contexts (example 3) but also in neutral way as a filler word in northern Italy (example 4).

7 Final Remarks

We explored the contribution of adding multilingual training material in the automatic identification of misogynist tweets in three languages: English, Spanish and Italian. Our models trained on monolingual data achieve state-of-the-art performance. The inclusion of data in one or two other languages impacts the performance negatively when compared to BERT models, but positively when compared to mBERT models. Multilingual models can be used as data augmentation technique — train on L_1+L_2 to predict on L_1, but they are not suitable for zero-shot classification across languages — train on L_1 to predict on L_2, hinting that misogyny might be language-specific, but further experiments are required.

8 Societal Impact and Limitations

This work represents a starting point toward investigating whether misogyny is language-specific. Analysing the differences of misogyny across languages and cultures is important, since it can help policymakers to develop country-specific policies to mitigate its impact. On Twitter, as well as on other platforms, interactions can be carried out in different languages. We head toward a real-world application, which considers the multilingualism of the platform. Users would benefit from a system able to flag misogynous tweets in multiple languages. Such system would raise awareness and ultimately make a more enjoyable online environment for women.

Among the limitations of this work, we currently focus on three languages only, neglecting geographical information. As a result, not enough attention is paid to culture. Moreover, currently our models are not interpretable, and that would be an important aspect to raise awareness in the general public.

References

Resham Ahluwalia, Evgeniia Shcherbinina, Edward Callow, Anderson Nascimento, and Martine De Cock. 2018. Detecting misogynous tweets. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018), co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)*, page 2150:242–248, Sevilla, Spain. CEUR-WS.org.

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.

Amir Bakarov. 2018. Vector space models for automatic misogyny identification. In *Proceedings of the sixth evaluation campaign of natural language processing and speech tools for Italian. Final workshop (EVALITA 2018) co-located with the fifth Italian conference on computational linguistics (clic-it 2018)*. CEUR-WS.org.

Angelo Basile and Chiara Rubagotti. 2018. Crotonemilano for AMI at evalita2018. A performant, cross-lingual misogyny detection system. In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Turin, Italy, December 12-13, 2018*, volume 2263 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Valerio Basile, Cristina Bosco, Elisabetta Fersini, Debora Nozza, Viviana Patti, Francisco Manuel Rangel Pardo, Paolo Rosso, and Manuela Sanguinetti. 2019. SemEval-2019 task 5: Multilingual detection of hate speech against immigrants and women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 54–63, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Khandis R. Blake, Siobhan M. O'Dean, James Lian, and Thomas F. Denson. 2021. Misogynistic tweets correlate with violence against women. *Psychological Science*, 32(3):315–325.

Davide Buscaldi. 2018. Tweetaneuse@AMI EVALITA2018: character-based models for the automatic misogyny identification task (short paper). *CEUR Workshop Proceedings: 2263. Proceedings of the sixth evaluation campaign of natural language processing and speech tools for italian. final workshop (EVALITA 2018) co-located with the fifth italian conference on computational linguistics (clic-it 2018), turin, italy, december 12-13, 2018 (pp. 1–4). CEUR-WS.org.*

Camilla Casula and Sara Tonelli. 2020. Hate speech detection with machine-translated data: The role of annotation scheme, class imbalance and undersampling. *Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020).*

José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2020. Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.

Adriano dos S. R. da Silva and Norton T. Roman. 2020. No place for hate speech@AMI: Convolutional neural network and word embedding for the identification of misogyny in italian (short paper). *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020).*

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, MN. ACL.

Samuel Fabrizi. 2020. fabsam@AMI: A convolutional neural network approach. In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020).*

Diane Felmlee, Paulina Inara Rodis, and Amy Zhang. 2020. Sexist Slurs: Reinforcing Feminine Stereotypes Online. *Sex Roles*, 83(1):16–28.

Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2018. Overview of the Evalita 2018 task on automatic misogyny identification (AMI). In *EVALITA Evaluation of NLP and Speech Tools for Italian: Proceedings of the Final Workshop 12-13 December 2018, Naples*, pages 59–66. Torino: Accademia University Press.

Elisabetta Fersini, Debora Nozza, and Paolo Rosso. 2020. AMI @ EVALITA2020: Automatic misogyny identification. In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR.org.

Simona Frenda, Bilal Ghanem, and Manuel Montes y Gómez. 2018. Exploration of misogyny in spanish and english tweets. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018), co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018).*

Iakes Goenaga, Aitziber Atutxa, Arantza Casillas, Arantza Díaz de Ilarraza, Nerea Ezeiza, Koldo Gojenola, Maite Oronoz, Alicia Pérez, and Olatz Perez de Viñaspre. 2018. Automatic misogyny identification using neural networks. *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018).*

Alyssa Lees, Jeffrey Scott Sorensen, and Ian D. Kivlichan. 2020. Jigsaw@AMI and haspeede2: Fine-tuning a pre-trained comment-domain bert model. In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020).*

Chi-Liang Liu, Tsung-Yuan Hsu, Yung-Sung Chuang, and Hung yi Lee. 2020. What makes multilingual bert multilingual? *arXiv*.

Ilya Loshchilov and Frank Hutter. 2017. Fixing weight decay regularization in adam. *CoRR*, abs/1711.05101.

Arianna Muti and Alberto Barrón-Cedeño. 2020. UniBO@AMI: A Multi-Class Approach to Misogyny and Aggressiveness Identification on Twitter Posts Using AlBERTo. In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*.

Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2020. Misogyny detection in twitter: a multilingual and cross-domain study. *Information Processing & Management*, 57.

Endang Wahyu Pamungkas, Alessandra Teresa Cignarella, Valerio Basile, and Viviana Patti. 2018. 14-ExLab@UniTo for AMI at IberEeal2018: Exploiting lexical knowledge for detecting misogyny in English and Spanish tweets. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018), co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018)*.

Endang Wahyu Pamungkas and Viviana Patti. 2019. Cross-domain and cross-lingual abusive language detection: A hybrid approach with deep learning and a multilingual lexicon. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 363–370, Florence, Italy. Association for Computational Linguistics.

Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. AlBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets. In *Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019)*, volume 2481, Bari, Italy. CEUR.

Punyajoy Saha, Binny Mathew, Pawan Goyal, and Animesh Mukherjee. 2018. Indian institute of engineering science and technology (shibpur), indian institute of technology (kharagpur). In *Proceedings of the Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2018) co-located with the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018) (pp.1-9)*.

Kalpana Srivastava, Suprakash Chaudhury, P.S. Bhat, and Samiksha. Sahu. 2017. Misogyny, feminism, and sexual harassment. *Industrial psychiatry journal*, 26(2):111–113.

Alice Tontodimamma, Eugenia Nissi, Annalina Sarra, and Lara Fontanella. 2021. Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126(1):157–179.

Rob van der Goot, Nikola Ljubešić, Ian Matroos, Malvina Nissim, and Barbara Plank. 2018. Bleaching text: Abstract features for cross-lingual gender prediction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 383–389, Melbourne, Australia. Association for Computational Linguistics.