Laboratoire
Linguistique, Langues, Parole | LILPA | UR 1339
Université de Strasbourg

GeNRe: A French Gender-Neutral Rewriting System Using Collective Nouns

Enzo Doyen & Amalia Todirascu

June 2025

📖 Outline

I. Purpose and Objectives

II. Methodology and Models

III. Results

I. Purpose and Objectives

French language and masculine generics

In French, grammatical gender matches the referent's for humans:

un dans<u>eur</u> (*a dancer*$^{\text{male}}$); une dans<u>euse</u> (*a dancer*$^{\text{female}}$)

French language and masculine generics

In French, grammatical gender matches the referent's for humans:

> un danseur (*a dancer*$^{\text{male}}$); une danseuse (*a dancer*$^{\text{female}}$)

Masculine gender considered as default (**masculine generics**; MG) when referring to groups of both men and women:

> « **Les étudiants**$^{\text{MG}}$ ont été invités à participer à l'atelier. »
> (***Students***^{MG} *were invited to participate to the workshop.*)

Masculine generics and bias

Psycholinguistics studies in French and other languages with masculine generics show they bias towards male-centric mental representations **(Stahlberg et al., 2001; Gygax et al., 2008; Gygax et al., 2012; Harris Interactive, 2017; Körner et al., 2022)**.

Masculine generics and bias

Psycholinguistics studies in French and other languages with masculine generics show they bias towards male-centric mental representations **(Stahlberg et al., 2001; Gygax et al., 2008; Gygax et al., 2012; Harris Interactive, 2017; Körner et al., 2022)**.

Why is this important for NLP?

- **in machine translation systems**, potential **gender bias** when translating from/to languages with different gender systems **(Prates et al., 2020; Savoldi et al., 2021; Vanmassenhove, 2024)**
- **in text generation models/LLMs**, increased male representation **at the expense of women and non-binary people** **(Doyen & Todirascu, 2025)**

Task of Gender Rewriting

Definition: generating one or more alternative sentences that either neutralize gender, adopt inclusive forms, or switch to a different gender

Task of Gender Rewriting

Definition: generating one or more alternative sentences that either neutralize gender, adopt inclusive forms, or switch to a different gender

- User-Centric (M → F / F → M) Rewriting (Arabic: **Alhafni et al. (2022a); (2022b); Habash et al. (2019)**);
- Inclusive Rewriting (German: **Pomerenke (2022)**; Portuguese: **Veloso et al. (2023)**; French: **Lerner et Grouin (2024)**);
- Neutral Rewriting (English: **Sun et al. (2021); Vanmassenhove et al. (2021)**).

Task of Gender Rewriting

Definition: generating one or more alternative sentences that either neutralize gender, adopt inclusive forms, or switch to a different gender

- User-Centric (M → F / F → M) Rewriting (Arabic: **Alhafni et al. (2022a); (2022b); Habash et al. (2019)**);
- Inclusive Rewriting (German: **Pomerenke (2022)**; Portuguese: **Veloso et al. (2023)**; French: **Lerner et Grouin (2024)**);
- Neutral Rewriting (English: **Sun et al. (2021); Vanmassenhove et al. (2021)**).

No French neutral rewriting system, and collective nouns unexploited for gender-neutralization in French

Neutral Rewriting for French

Our work uses **human collective nouns** **(Lecolle, 2019)** for neutralizing gender, as their grammatical gender does not depend upon the referent's.

Examples: personnel (*staff*), jury, police, armée (*army*)...

Masculine generics		**Collective noun**
Tous les anciens **employés**^{MG} sont invités au gala. (*All former* ***employees***^{MG} *are invited to the gala*.)	→	L'ancien **personnel** est invité au gala. (*Former* ***staff*** *is invited to the gala*.)

Neutral Rewriting for French

Our work uses **human collective nouns** **(Lecolle, 2019)** for neutralizing gender, as their grammatical gender does not depend upon the referent's.

Examples: personnel (*staff*), jury, police, armée (*army*)...

Masculine generics		**Collective noun**
Tous les anciens **employés**^{MG} sont invités au gala. (*All former* ***employees***^{MG} *are invited to the gala*.)	→	L'ancien **personnel** est invité au gala. (*Former* ***staff*** *is invited to the gala*.)

Neutral Rewriting for French

Our work uses **human collective nouns** **(Lecolle, 2019)** for neutralizing gender, as their grammatical gender does not depend upon the referent's.

Examples: personnel (*staff*), jury, police, armée (*army*)...

Masculine generics

Tous les anciens **employés**^{MG} sont invités au gala.

(*All former* ***employees***^{MG} *are invited to the gala.*)

→

Collective noun

L'ancien **personnel** est invité au gala.

DET ADJ VERB

(*Ø Former* ***staff*** *is invited to the gala.*)

II. Methodology and Models

Methodology

1. Creating a dictionary with French collective nouns and their masculine generics member nouns counterparts, gathering **315 entries**.

Member noun (MG)	Collective noun
employés (*employees*)	personnel (*staff*)
jurés (*jury members*)	jury
policiers (*police officers*)	police
soldats (*soldiers*)	armée (*army*)
…	…

Methodology

1. Creating a dictionary with French collective nouns and their masculine generics member nouns counterparts, gathering **315 entries**.

Member noun (MG)	Collective noun
employés (*employees*)	personnel (*staff*)
jurés (*jury members*)	jury
policiers (*police officers*)	police
soldats (*soldiers*)	armée (*army*)
…	…

2. Extracting sentences containing member nouns from a French Wikipedia corpus **(graelo, 2023)** and Europarl **(Koehn, 2005)**, got **398,954 sentences**.

Methodology

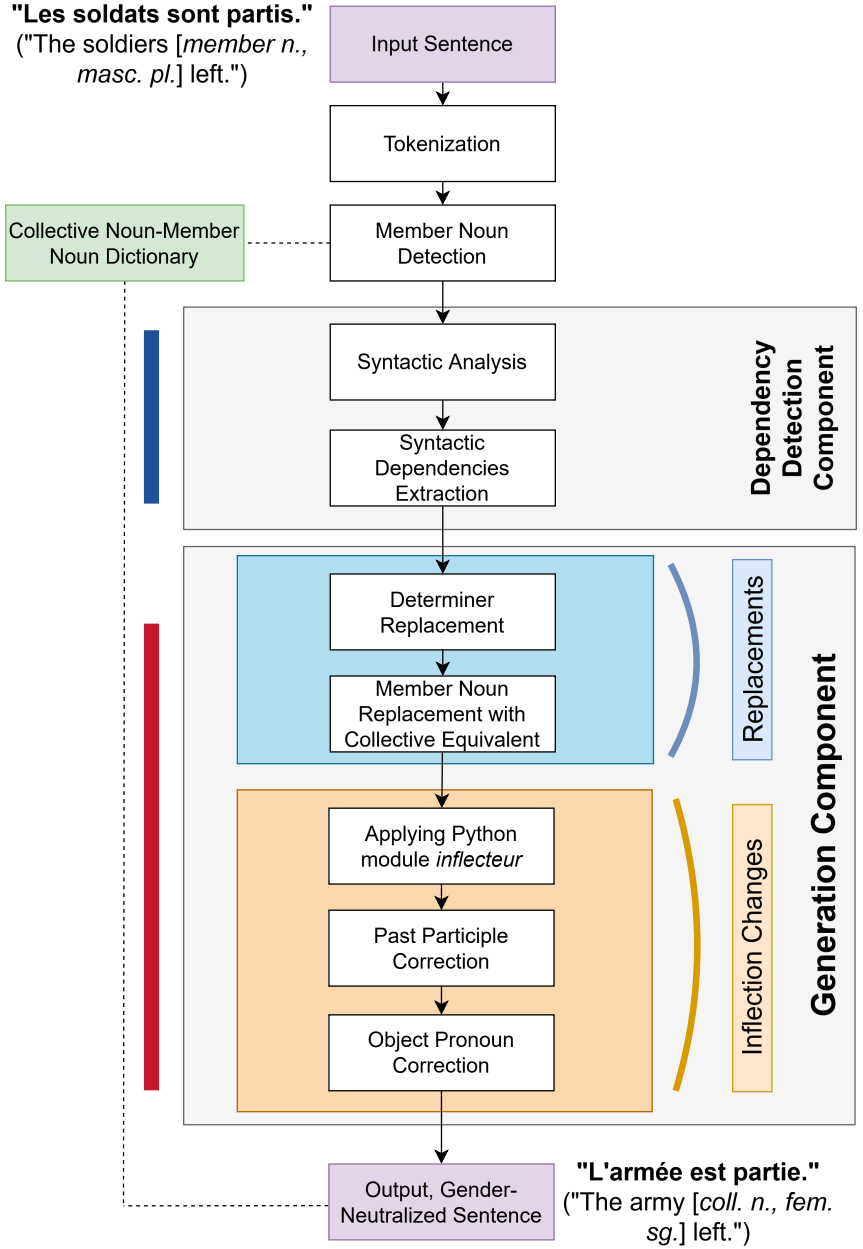
1. Creating a dictionary with French collective nouns and their masculine generics member nouns counterparts, gathering **315 entries**.

Member noun (MG)	Collective noun
employés (*employees*)	personnel (*staff*)
jurés (*jury members*)	jury
policiers (*police officers*)	police
soldats (*soldiers*)	armée (*army*)
...	...

2. Extracting sentences containing member nouns from a French Wikipedia corpus **(graelo, 2023)** and Europarl **(Koehn, 2005)**, got **398,954 sentences**.
3. Developing **three different models for rewriting**: a rule-based system (RBS), fine-tuned models and an instruct-based model.

Rule-based system (RBS)

- Tokenizing input with *spaCy* v.3.7.4 **(Montani et al., 2024)**, detecting member nouns using our member noun-collective noun dictionary
- Using **two main components**:
  - **dependency detection component** to find member noun syntactic dependencies (determiners, adjectives, verbs…);
  - **generation component** to replace member nouns with their collective equivalents and reinflect using Python library *inflecteur* **(Chuttarsing, 2021)**: **73.42% acc.** on our data
    - additional changes by our RBS: **+2.34 acc.**



Fine-tuned models

- Finetuned T5-small **(Raffel et al., 2020)** & M2M100 **(Fan et al., 2020)** on 60,000 RBS-converted sentences (Wikipedia/Europarl). Used 6,000 (10%) for eval. Hypothetical 0.27 WER improv. shown by **Vanmassenhove et al. (2021)**.

Fine-tuned models

- Finetuned T5-small **(Raffel et al., 2020)** & M2M100 **(Fan et al., 2020)** on 60,000 RBS-converted sentences (Wikipedia/Europarl). Used 6,000 (10%) for eval. Hypothetical 0.27 WER improv. shown by **Vanmassenhove et al. (2021)**.

Instruct-based model

- Model type not used in other works for this specific task. Using Claude 3 Opus **(Anthropic, 2024)** with few-shot prompting.

Instruction Type	Description
BASE	Asking to replace member nouns with their collective noun equivalents, without specifying replacements
DICT	Asking to replace member nouns with their collective noun equivalents, specifying replacements using the dictionary
CORR	Asking to correct RBS-converted sentence

III. Results

Dependency Detection Results (RBS)

- Manually annotated dependencies from 500 sentences (250 Wikipedia; 250 Europarl)
- **Objective**: evaluate correct detection of syntactic dependencies of the member noun to be modified
- **Baseline**: default spaCy-detected member noun's dependencies (excluding punctuation)

	Wikipedia			Europarl			Avg.		
	Precision	**Recall**	**F1**	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Baseline	0.096	0.723	0.169	0.115	0.689	0.197	0.1055	0.706	0.183
GeNRe-RBS	0.773	0.855	0.812	0.758	0.813	0.785	0.7655	0.834	0.7985

Generation Results (RBS/fine-tuned/instruct-based)

- Manually gender-neutralized 500 sentences (250 Wikipedia; 250 Europarl) using collective nouns
- **Objective**: evaluate correct conversion of sentences
- **Baseline**: original, unconverted sentence (as per previous works)

Generation Results (RBS/fine-tuned/instruct-based)

- Manually gender-neutralized 500 sentences (250 Wikipedia; 250 Europarl) using collective nouns
- **Objective**: evaluate correct conversion of sentences
- **Baseline**: original, unconverted sentence (as per previous works)

Metrics:

- **WER** (word difference between reference and converted sentence)
- **BLEU** (n-gram overlap between reference and converted sentence, $n = 4$)
- **cosine similarity** (sentence-level embeddings comparison with SBERT **(Reimers & Gurevych, 2019)** and model `Sentence-CamemBERT-Large`)

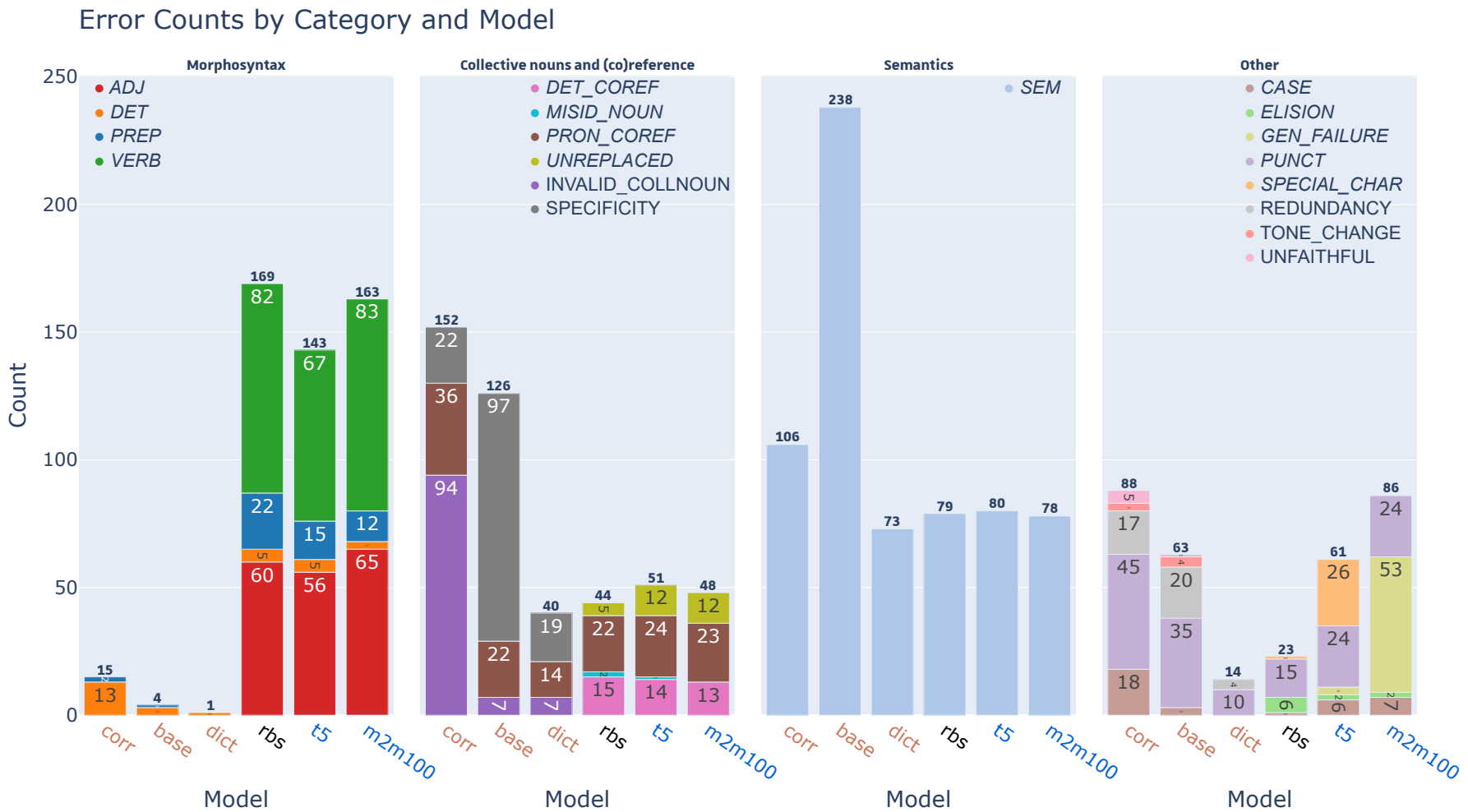
Generation Results

System	WER (↓)	BLEU (↑)
GeNRe-RBS (ours)	-8,449% **(avg.)**	+11,10 **(avg.)**
Sun et al. (2021, English)	-11,77%	+9,63
Vanmassenhove et al. (2021, English)	-10,947 % (avg.)	—

	Wikipedia			Europarl			Avg.		
	WER (↓)	**BLEU (↑)**	**Cosine (↑)**	**WER (↓)**	**BLEU (↑)**	**Cosine (↑)**	**WER (↓)**	**BLEU (↑)**	**Cosine (↑)**
Baseline	12.611%	80.688	97.436	12.446%	82.871	97.008	12.529%	81.779	97.222
GeNRe-RBS	**4.03%**	92.096	**98.88**	**3.814%**	93.707	99.22	**3.81%**	92.887	**99.05**
GeNRe-T5	6.726%	87.358	98.508	4.259%	93.111	99.1	5.492%	90.234	98.804
GeNRe-M2M100	6.566%	88.186	97.232	4.247%	93.197	98.992	5.406%	90.692	98.112
Claude-BASE	13.87%	80.205	96.532	10.713%	85.313	97.128	12.291%	82.759	96.83
Claude-DICT	4.702%	**92.79**	98.812	4.197%	**94.247**	**99.264**	4.45%	**93.519**	99.038
Claude-CORR	11.282%	84.954	98.092	8.992%	85.257	98.056	10.137%	85.25	98.074

RBS/Baseline difference comparison with other gender-neutralization systems (**Table 1**) and Global results (**Table 2**)

Error Counts and Types



Conclusion

Contributions

- Dictionary of French collective nouns and their member noun equivalents
- Dataset of gender-neutralized and non-gender-neutralized sentences
- First automatic gender-neutralization system for French
- Highlighted potentiality of instruct-based models combined with pre-created resources for this task

Limits

- Specific semantics of collective nouns (making them unusable in some contexts)
- Limits of LLM-as-a-judge approach for instruct-based model error analysis **(Gu et al., 2025)**

Our work will be presented in-person in Vienna.
Looking forward to having a chat there!

Bibliography

Alhafni, B., Habash, N., & Bouamor, H. (2022). User-Centric Gender Rewriting. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 618–631. https://doi.org/10.18653/v1/2022.naacl-main.46

Alhafni, B., Habash, N., Bouamor, H., Obeid, O., Alrowili, S., Alzeer, D., Alshanqiti, K. M., ElBakry, A., ElNokrashy, M., Gabr, M., Issam, A., Qaddoumi, A., Vijay-Shanker, K., & Zyate, M. (2022). *The Shared Task on Gender Rewriting*. https://doi.org/10.48550/arXiv.2210.12410

Anthropic. (March 2024). *The Claude 3 Model Family: Opus, Sonnet, Haiku*.

Chuttarsing, A. (2021). *Inflecteur*.

Doyen, E., & Todirascu, A. (14 February 2025). *Man Made Language Models? Evaluating LLMs' Perpetuation of Masculine Generics Bias*. https://doi.org/10.48550/arXiv.2502.10577

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., Baines, M., Celebi, O., Wenzek, G., Chaudhary, V., Goyal, N., Birch, T., Liptchinsky, V., Edunov, S., Grave, E., Auli, M., & Joulin, A. (October 2020). *Beyond English-Centric Multilingual Machine Translation*.

graelo. (2023). *Graelo/Wikipedia*.

Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Wang, Y., Gao, W., Ni, L., & Guo, J. (2025). *A Survey on LLM-as-a-Judge*. https://arxiv.org/abs/2411.15594

Gygax, P., Gabriel, U., Lévy, A., Pool, E., Grivel, M., & Pedrazzini, E. (June 2012). The Masculine Form and Its Competing Interpretations in French: When Linking Grammatically Masculine Role Names to Female Referents Is Difficult. *Journal of Cognitive Psychology*, *24*(4), 395–408. https://doi.org/10.1080/20445911.2011.642858

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (April 2008). Generically Intended, but Specifically Interpreted: When Beauticians, Musicians, and Mechanics Are All Men. *Language and Cognitive Processes*, *23*(3), 464–485. https://doi.org/10.1080/01690960701702035

Habash, N., Bouamor, H., & Chung, C. (2019). Automatic Gender Identification and Reinflection in Arabic. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, 155–165. https://doi.org/10.18653/v1/W19-3822

Harris Interactive. (2017). *L'écriture Inclusive : La Population Fraņcaise Connaît-Elle l'écriture Inclusive ? Quelle Opinion En a-t-Elle ?*.

Koehn, P. (2005). *Europarl: A Parallel Corpus for Statistical Machine Translation*.

Körner, A., Abraham, B., Rummer, R., & Strack, F. (October 2022). Gender Representations Elicited by the Gender Star Form. *Journal of Language and Social Psychology*, *41*(5), 553–571. https://doi.org/10.1177/0261927X221080181

Lecolle, M. (2019). *Les Noms Collectifs Humains En Fraņcais. Enjeux Sémantiques, Lexicaux et Discursifs* (C. Jacquet-Pfau, Ed.). Lambert-Lucas.

Lerner, P., & Grouin, C. (2024). *INCLURE: A Dataset and Toolkit for Inclusive French Translation*.

Montani, I., Honnibal, M., Boyd, A., Landeghem, S. V., & Peters, H. (October 2024). *spaCy: Industrial-strength Natural Language Processing in Python*. https://doi.org/10.5281/ZENODO.1212303

Pomerenke, D. (December 2022). *INCLUSIFY: A Benchmark and a Model for Gender-Inclusive German*.

Prates, M. O. R., Avelar, P. H., & Lamb, L. C. (May 2020). Assessing Gender Bias in Machine Translation: A Case Study with Google Translate. *Neural Computing and Applications*, *32*(10), 6363–6381. https://doi.org/10.1007/s00521-019-04144-6

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (July 2020). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. http://arxiv.org/abs/1908.10084

Savoldi, B., Gaido, M., Bentivogli, L., Negri, M., & Turchi, M. (August 2021). Gender Bias in Machine Translation. *Transactions of the Association for Computational Linguistics*, *9*, 845–874. https://doi.org/10.1162/tacl_a_00401

Stahlberg, D., Sczesny, S., & Braun, F. (December 2001). Name Your Favorite Musician: Effects of Masculine Generics and of Their Alternatives in German. *Journal of Language and Social Psychology*, *20*(4), 464–469. https://doi.org/10.1177/0261927X01020004004

Sun, T., Webster, K., Shah, A., Wang, W. Y., & Johnson, M. (February 2021). *They, Them, Theirs: Rewriting with Gender-Neutral English* (Issue arXiv:2102.06788). arXiv.

Vanmassenhove, E. (January 2024). *Gender Bias in Machine Translation and The Era of Large Language Models* (Issue arXiv:2401.10016). arXiv. https://doi.org/10.48550/arXiv.2401.10016

Vanmassenhove, E., Emmery, C., & Shterionov, D. (2021). *NeuTral Rewriter: A Rule-Based and Neural Approach to Automatic Rewriting into Gender-Neutral Alternatives*. https://doi.org/10.48550/arXiv.2109.06105

Veloso, L., Coheur, L., & Ribeiro, R. (2023). A Rewriting Approach for Gender Inclusivity in Portuguese. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 8747–8759. https://doi.org/10.18653/v1/2023.findings-emnlp.585