

**CLiC-it 2023
Italian Conference
on Computational Linguistics**

**Proceedings
of the
9th Italian Conference
on
Computational Linguistics**

Venice, Italy, November 30 - December 2, 2023

Edited by:

**Federico Boschetti
Gianluca E. Lebani
Bernardo Magnini
Nicole Novielli**



aA



Beyond Obscuration and Visibility: Thoughts on the Different Strategies of Gender-Fair Language in Italian

Martina Rosola¹, Simona Frenda^{2,3}, Alessandra Teresa Cignarella^{2,3}, Matteo Pellegrini⁴, Andrea Marra² and Mara Floris⁵

¹Università degli Studi di Brescia, Italy

²Department of Computer Science, University of Turin, Italy

³aequa-tech, Turin, Italy

⁴CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan, Italy

⁵Università Vita-Salute San Raffaele, Milan, Italy

Abstract

This study focuses on the growing importance of gender-fair language and explores innovative strategies proposed also in other languages to avoid gender-specific endings. We present a set of guidelines for the annotation and reformulation of gender-(un)fair texts and their application to a corpus of 1,024 portions of university administrative documents in Italian. Overall, the guidelines presented in this study prove to be valuable both practically and theoretically. They help identify and address non-inclusive expressions while highlighting the complexities of *obscuration* and *visibility* in gender-fair language reformulation. In addition, the statistical analysis of the created corpus shows how administrative texts tend to contain gender-unfair language, especially the masculine overextended expressions, showing the need to adopt specific and complete guidelines that lead (and support the staff training) to the use of a more gender-fair language.

Keywords

Annotation Schema, Italian, Gender-fair language

1. Introduction

Gender-fair language (GFL), also known as inclusive language, refers to the use of linguistic expressions that promote gender equality and avoid reinforcing gender stereotypes. The Italian language, like many others, has traditionally exhibited gender asymmetries and biases, which have perpetuated gender inequality and exclusion. However, in recent years, there has been a growing awareness and effort to address these issues by promoting GFL. In this work, we present the annotation scheme we developed to identify and reformulate gender-unfair expressions, and the corpus we applied it to, encompassing a range of administrative documents of the University of Brescia¹.

The significance of addressing gender-unfair expressions stems from concerns raised by several researchers. In Italian, a personal noun's grammatical gender typically correspond to its referent's gender. In certain cases, however, a discrepancy between the two arises. Crucially, such discrepancies are characterized by an asymmetrical

nature, as has been highlighted both from a theoretical and a practical perspective [1, 2]. In Italian, as well as in other languages with gendered nouns, the issue of GFL presents a dual challenge. Firstly, the binary distinction between masculine and feminine in the Italian gender system excludes individuals who identify outside the male-female dichotomy. Consequently, non-binary individuals are consistently misgendered due to the absence of dedicated linguistic forms. Secondly, the usage of the generic masculine (e.g., using masculine profession names to refer to individuals of any gender) and the overextended masculine (referring to mixed-gender groups using the masculine plural) predominantly evoke masculine mental representations, thereby limiting the visibility of female and non-binary individuals. Notably, the presence of a single man in a group is sufficient to alter the gender of the term used to refer to the group, whereas the reverse does not hold true for inverted genders.

These discrepancies, in addition to their asymmetrical nature, impact the mental representations we construct. Despite the Italian convention of using masculine terms to refer to individuals of unknown gender and mixed groups, psycholinguistic studies highlight potential issues in this respect. Extensive experimentation conducted over several years using various techniques, and across different languages, suggests that the overextended masculine and generic masculine are interpreted as if they were purely masculine [e.g., 3, 4].

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 – Dec 02, 2023, Venice, Italy

✉ martina.rosola@unibs.it (M. Rosola); simona.frenda@unito.it (S. Frenda); alessandrateresa.cignarella@unito.it (A. T. Cignarella); matteo.pellegrini@unicatt.it (M. Pellegrini); andrea.marra@unito.it (A. Marra); floris.mara@hsr.it (M. Floris)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹<https://www.unibs.it>

Recognizing the importance of rectifying such language discrepancies, various guidelines have been published over the years [e.g., 5]. The annotation scheme we propose draws upon the recommendations presented in the available guidelines, to develop a comprehensive framework for addressing gender-unfair expressions in Italian language usage. To the best of our knowledge, our annotation scheme represents a novel approach. While another project (i.e., E-MIMIC) focuses on inclusive language, it simply distinguishes between inclusive and non-inclusive texts [6]. Our annotation scheme appears to be the first one distinguishing between different types of gender-unfair language, and it comprehensively considers all the gender-fair options when it comes to offering alternative wordings. Moreover, applying this scheme of annotation to various administrative texts, we showed how, despite the existence of various guidelines, they remain impregnated with gender-unfair expressions.

In this work, firstly we review previous studies on this topic, both in theoretical linguistics (subsection 2.1) and in NLP (subsection 2.2). We then describe in detail the annotation scheme (section 3) and the creation of the annotated corpus (section 4), also providing a preliminary analysis of the data gathered so far.

2. Related Work

2.1. Linguistics

Sexism in language and how to make Italian gender-fair are increasingly studied and debated topics (see [7] for an overview). The classic reference point in the literature is Sabatini [8], which comprises an analysis of sexism in the Italian language and recommendations on how to overcome it.

Sabatini [8] identifies *grammatical* and *semantic* asymmetries, namely gender-unfair grammatical and discursive or lexical linguistic conventions. The use of masculine terms for mixed-gender groups belongs to the former, while the exclusive use of adjectives for one gender (e.g., GRAZIOSO, ‘pretty’, is hardly used for men) instantiates the latter. On top of avoiding semantically sexist expressions, Sabatini [8] advises using feminine job titles for women and conjoining masculine and feminine forms for mixed-gender groups.

Her recommendations have been expanded and adapted by several private and public bodies, that issued gender-fair language guidelines [e.g., 9, 5, 10]. These works, among others, distinguish between strategies aimed at symmetrizing language by giving women the same visibility that men have, and strategies aimed at getting rid of sexism through the avoidance of gendered forms altogether. While Sabatini and the subsequent guidelines focus on the discrimination and

(in)visibilization of *women* in language, more recent scholarly and activist debates also concern how to address and talk about *non-binary people*, i.e., those that do not exclusively identify as men or women, aiming at making them visible too.

The Italian grammatical gender system, indeed, is binary and does not provide a straightforward way to refer to non-binary people. Various gender-neutral suffixes are in use, such as ‘-ə’ or ‘-u’ (see [11] for a comprehensive list). As González Vázquez et al. [12] observe, such *innovative* proposals can be employed to make gender visible, as in “tutti, tutte e tutto” (everyone:M.PL, everyone:F.PL, and everyone:INN.PL)², or to neutralize it, as in “tuttu” (everyone:INN.PL) used for a mixed-gender group.

The implementation of innovative strategies also depends on the features of the language. Marcato and Thüne [13] provide an analysis of the Italian grammatical gender system, distinguishing between nouns whose referential gender is expressed by different lexical roots (e.g., “madre”, mother:F.SG, and “padre”, father:M.SG); nouns with *mobile* gender, whose referential gender is specified through the addition of different suffixes to the same lexical root (e.g., “figlia”, daughter:F.SG, and “figlio”, son:M.SG); and the so-called *epicene* nouns, whose gender is not overtly marked, but only revealed by satellite elements – i.e., the noun’s determiners and modifiers (e.g., “la nipote”, the.F.SG niece:F.SG; “il nipote”, the.M.SG nephew:M.SG). As Formato [14] observes, some nouns (i.e., ‘semi-epicene’) work in the latter way only in the singular and have different gendered suffixes in the plural (e.g., “giornalista”, journalist:F/M.SG; “giornaliste”, journalist:F.PL; “giornalisti”, journalist:M.PL). Finally, a few nouns refer to individuals of any gender irrespective of their grammatical gender (e.g., “persona”, person:F.SG).

Due to this peculiar characteristic, these nouns can be straightforwardly used to refer to non-binary people as well. Moreover, gender-neutral suffixes are not required for epicene (and, in the singular, semi-epicene) nouns, as they are not overtly marked for gender. In this case, the only needed precaution to get a gender-neutral form concerns the choice of gender-neutral satellite elements or their gender-neutralization. Gender-neutral suffixes are further ineffective for nouns like MADRE and PADRE, where it is the root to be overtly marked for gender. These word endings, thus, should only be used with nouns with mobile gender, which, however, constitute the vast majority of Italian animate nouns (see [15], p. 106). Innovative strategies should also be used for the many gendered pronouns, determiners, past participles, and adjectives in order to make them gender-neutral and suitable to refer to non-binary people and to mixed-gender groups.

Formato [14] also provides a taxonomy of linguistic us-

²We label such forms as INN ‘innovative’.

ages influenced by gendered assumptions. Like Sabatini, Formato focuses on both sexist expressions and linguistic conventions. Among the latter, Formato originally distinguishes the case in which masculine terms are used for mixed-gender groups from those in which they are used for unknown or generic individuals.

In our framework, we elaborate on the categories identified in these works, to develop our own taxonomy, that we implement in the annotation scheme described in section 3.

2.2. Natural Language Processing

In recent years, sexism and gender-(un)fair practices have been addressed in Computational Linguistics, mostly focusing on the presence of gender bias in automatic systems. As highlighted in Costa-jussà [16], studies on gender bias in NLP serve a dual role. On the one hand, NLP can function as a tool to identify gender bias in various social domains such as online news or advertisements. On the other hand, it frequently generates gender-biased systems, thus contributing to the perpetuation and reinforcement of gender bias within society. This bias in NLP is predominantly attributed to the training of models on datasets that exhibit inherent biases. Consequently, the amplification of bias occurs through the learning algorithms employed in NLP systems.

Some specific studies have been conducted in the field of Machine Translation. One of the most recent was carried out by Rescigno et al. [17] who explored how three of the most popular translation systems (Google Translate, Bing Microsoft Translator, and DeepL) handle gender phenomena in natural languages, such as pronouns, job titles, and occupation names. The authors compared the translations generated from English respectively to Italian, French, and Spanish, revealing that all three systems exhibit some level of gender bias, with Google Translate producing more biased translations, Bing Microsoft Translator displaying a lesser degree of bias, and DeepL generally being more gender-neutral.

Similarly to Costa-jussà [16], Sun et al. [18] conducted a comprehensive literature review, exploring various strategies proposed in existing research to address gender bias, including dataset preprocessing, algorithmic modifications, and post-processing techniques. The paper emphasizes the significance of mitigating gender bias in NLP systems and highlights the challenges associated with bias detection and mitigation.

More recently, Stanczak and Augenstein [19] identify four key limitations in current research on gender bias in NLP. Firstly, social gender is often treated as a binary variable, not paying attention to its fluidity and continuity. Secondly, studies usually give more importance to high-resource languages - primarily English - neglecting the diversity of languages spoken globally. Thirdly, despite

the abundance of papers on gender bias, many newly developed algorithms lack sufficient bias testing and fail to address ethical considerations. Lastly, the methodologies employed in this area often lack comprehensive definitions of gender bias and robust evaluation baselines and pipelines.

The present work contributes to address many of these issues: we explicitly take into account the linguistic representation of non-binary individuals, we create an annotation scheme for Italian – for which much fewer NLP studies are available compared to English – and finally, we present an annotated corpus that could be exploited for the training of automatic NLP tools.

3. Annotation Scheme

The annotation task is divided into two parts. A first annotation layer concerns the identification of portions of text(s) where gender-unfair language is used, and the assignment of each of them to a specific type among the following ones:

- ‘incongruous’ (It. “incongruo”), when the grammatical gender of the noun (and, possibly, of its modifiers), does not match the gender of the referent identified in discourse (e.g., “il ministro del turismo, Daniela Santanché”, the.M.SG minister:M.SG of.the tourism, Daniela Santanché);
- ‘overextended’ (It. “sovraesteso”), when the masculine (or, in rare cases, feminine) grammatical gender is used to refer to a group composed of people with different genders (e.g., “il rapporto con i professori è buono”, the relationship with the.M.PL professor:M.PL is good, with reference to a group of teachers possibly comprising men, women, and non-binary individuals);
- ‘generic’ (It. “generico”) when the masculine (or, in rare cases, feminine) grammatical gender is used to refer to a generic or specific, but unknown, person, whose actual gender cannot be guessed (e.g. “il vincitore riceverà un premio”, the.M.SG winner:M.SG will.receive a bonus, where the identity of the winner is unknown, and so is their gender).

A second annotation layer concerns the proposal of gender-fair reformulations of the portions of texts identified as unfair,³ and the assignment of each reformulation to a specific type.

As for cases of incongruous gender, the only possible type of gender-fair solution is providing a ‘congruous’ (It. “congrua”) alternative option, where the grammatical gender matches the gender of the referent (e.g., “la

³At least one reformulation for each gender-unfair portion is required, but more than one reformulation is often provided.

	visibility	obscuration
conservative	<i>il vincitore o la vincitrice</i>	<i>chi vincerà</i>
innovative	<i>il vincitore o la vincitrice o l* vincitor*</i>	<i>l* vincitor*</i>

Table 1
Four-way contrast of reformulation types

ministra del turismo, Daniela Santanché”, the.F.SG minister:F.SG of the tourism, Daniela Santanché). On the other hand, to classify the reformulations proposed for cases of overextended and generic uses of grammatical gender, we start from two orthogonal binary distinctions, namely:

- ‘visibility’ (It. “visibilità”) strategies, that make the possible reference to persons with different genders explicit by means of the use of different grammatical genders; vs. ‘obscuration’ (“oscuramento”) strategies, that try to avoid the use of expressions that reveal the (assumed) gender of referents;
- ‘conservative’ (It. “conservativa”) strategies, that only use expressions that are part of the grammatical system of the standard variety of Italian; vs. ‘innovative’ (It. “innovativa”) strategies, that introduce new means of expression into the system.

These distinctions generate a four-way contrast, that is illustrated in Table 1, where one example reformulation per type is provided for the phrase *il vincitore* (see the Appendix for other examples).

Lastly, ‘mixed’ (It. “ibride”) reformulations use different strategies for different elements in the gender-unfair portion of text, e.g. *l* vincitore o vincitrice* (the.INN.SG winner:M.SG or winner:F.SG), where an innovative obscuration strategy is used for the article and a conservative visibility strategy is used for the noun.

4. Corpus

The scheme of annotation was applied by 5 expert annotators of gender-fair language to a small corpus of administrative texts coming from the University of Brescia. Differently from other textual genres, the administrative texts, for their format and technical language, have required a specific preprocessing process to let the annotators focus especially on the spans of text that could contain discrepancies.

To this purpose, we employed the original lexicon of professional names taken into account in Sabatini [8], enriching it with terms especially pertaining to Academia or terms that could be used in an overextended way (i.e., “essi”, they:M.PL). Below, we describe the steps of documents collection, preprocessing, and annotation. Finally,

we also present some first analyses of the resulting corpus.

Data Collection The data made available by the University of Brescia include a range of administrative materials such as the department’s strategic plan, reports from the departmental council and parity commission, as well as various forms. Most of them are already public on the website of the University, others, like the forms, have been asked to administrative organs. For this pioneering study, we collected specifically 13 documents.

Data Preprocessing All the documents have been transformed into plain text to be processed automatically. To deal with the special format of forms or the layout in tables of special reports, we designed various regular expressions to clean and prepare the texts for the segmentation in sentences.

To support the annotation phase, we signaled for each sentence the possible presence of discrepancies: displayed below each task were any words from the enriched lexicon of professions’ names based on Sabatini [8] present in the sentence. From a total of 1,024 sentences, 409 contained such words. However, the sentences in which the annotators have detected at least one unfair expression are 422. The lexicon has been updated to include all the words pointed by the annotators.

Annotation Process All annotators were trained on the annotation scheme, which was analysed during an initial meeting. Doubtful cases were discussed in regular bi-weekly meetings. In addition, we kept a file of notes in which we reviewed and discussed uncertain cases as they arose. The annotation scheme was subsequently updated in response to the insights gathered both in the file and the meetings. The annotation process has been carried out on LabelStudio platform⁴, creating a specific interface that facilitates the two layers of annotation: the identification of the gender-unfair expression, and the reformulation with one or more alternatives. The interface provided a section for comments in order to encourage reflection on the annotation scheme and collect insights from the annotators.

Even if the amount of analyzed data seems small, the annotation task has been conducted from October 2022

⁴<https://labelstud.io/>

to January 2023 by 5 experts of gender-(un)fair language (philosophers of language, linguists, and computational linguists), that provided alternatives for each textual span identified as unfair in the sentence.

Preliminary Analysis Thanks to this process of annotation in two layers, we created a corpus of 422 sentences where at least one gender-unfair expression has been identified, and 602 sentences where no gender-unfair expression has been identified.

In the 422 sentences containing gender-unfair expressions, the annotators detected on average 3 textual spans per sentence containing gender-unfair language (for a total of 3,195 portions) and proposed from 1 to about 11 alternatives.

Moreover, looking at the frequencies of the types of unfair expressions identified in the corpus, we can see from Table 2 that the most common case of gender-unfair language in administrative documents is represented by the use of overextended forms, and in particular of overextended masculine (e.g., “i ricercatori”, the.M.PL researchers:M.PL; and “i docenti”, the.M.PL teachers, for mixed-gender groups of, respectively, researchers and teachers).

#_portions	type
2,709	overextended
452	generic
34	incongruous
3,195	total

Table 2

Frequencies of the portions of texts identified as containing gender-unfair language. The use of *generic* and *overextended* gender is mostly referred to masculine cases.

Agreement A quantitative measure of inter-annotator reliability has not been calculated for different reasons. First, the scheme provides a variety of gender-fair options and the choice of a specific alternative depends on several factors, including individual preference: one annotator might agree on “lə Professorə” (the Professor:INN.SG) being gender-fair, but choose a different innovative option, like “lu Professoru” (the Professor:INN.SG), instead. Therefore, comparing the alternatives provided by each annotator is not a good measure of whether the annotators consider a specific option an appropriate gender-fair alternative to a certain gender-unfair expression. Relatedly, we do not plan to release an aggregated dataset, inclusive of a set of “gold-standard” preferred labels. Indeed, the aim of the present work is not to consolidate a ‘ground truth’ among rephrasing strategies but rather to explore as many solutions as possible while using gender-fair language. This is tied to our focus on methodology:

the main purpose of this paper is to present the process of our work. As we mentioned before, we created a novel annotation scheme for the Italian language, which allows a fine-grained distinction between different cases of gender discrepancies. Moreover, the scheme has continuously been discussed between authors and annotators, mostly concerning the interpretation of labels, such as e.g. the sensible distinction between “overextended” and “generic” gender-unfair expressions. Last but not least, the identified span did not just contain gender-unfair expressions, but any element that needs to be changed in order to get a gender-fair text. For example, if an annotator decided to propose “il corpo docenti” (the teaching staff) instead of “i docenti” (the.M.PL teachers) in “i docenti devono partecipare” (the teachers have to participate), they also have to select the verb “devono” for it to agree in number with “il corpo docenti”. Crucially, the verb only has to be selected if the proposed alternative to “i docenti” is singular and, thus, the verb needs to be singular too. Hence, if another annotator doesn’t propose a singular alternative to “i docenti”, they won’t need to select the verb. As a result, the two annotators would select different spans even when agreeing on what are the gender-unfair expressions within the text. For this reason, even comparing just the textual spans between annotators would not be a good indicator of the annotators’ agreement.

However, in Table 3 below we provide an example of a sentence with the reformulations proposed by the five annotators for each gender-unfair span of text, to give an idea of the kind of variation that can be found.

5. Conclusions and Future Work

In recent years, gender-fair language has gained significant attention, leading to the proposal of new strategies in various languages to avoid using masculine or feminine endings. Motivated by these theories, we conducted a study to examine the usage of different solutions in practical situations. We developed guidelines for gender-fair annotation and reformulation of texts, which we applied to a corpus of 1,024 portions of university administrative documents in Italian.

The corpus was annotated by 5 experts, and in 422 cases the annotators identified at least one gender-unfair expression. The preliminary analysis of this corpus highlighted the need to adopt specific guidelines (as well as a list of words to pay particular attention to) to support administrative staff in writing gender-fair texts.

Applying our annotation and reformulation guidelines to real data has led to theoretical advancements: we discovered that ‘obscurity’ and ‘visibility’ strategies can coexist within the same reformulation, and we consequently updated the annotation scheme to include ‘mixed’

Text: A partire dal 2013 il DiGi ha organizzato ogni anno International Summer schools, allo scopo di attrarre studenti stranieri e di offrire agli studenti bresciani l'opportunità di entrare in contatto con studenti e docenti di altri Paesi.

ann.	span	Conservative visibility	Innovative visibility	Conservative obscuration	Innovative obscuration
A	'stranieri'	–	–	'di origine straniera' or 'di nazionalità estera'	–
	'agli studenti bresciani'	–	–	a 'studenti provenienti dalla provincia di Brescia'	–
B	'studenti stranieri'	–	–	'studenti di università estere'	–
	'agli studenti bresciani'	–	–	'a coloro che studiano all'Università di Brescia'	–
C	'studenti stranieri'	'studenti/esse stranieri/e'	'studenti/esse/ə stranieri/e/ə'–		'studentə straniera'
	'studenti bresciani'	'studenti/esse bresciani/e'	'studenti/esse/ə bresciani/e/ə'	–	'studentə bresciana'
	'studenti'	'studenti/esse'	–	'studentə'	–
D	'stranieri'	'straniere/i'	–	'dall'estero'	–
	'agli studenti'	'alle/agli studenti'	–	–	–
	'bresciani'	'bresciane/i'	–	'dalla provincia di Brescia'	–
E	'stranieri'	'straniere/i'	'straniere/i/3' or 'straniere/i/u'	'di nazionalità straniera'	–
	'agli studenti'	'alle/agli studenti'	'alle/agli/all3 studenti' or 'alle/agli/allu studenti'	'alle persone che studiano'	'all3 studenti' or 'all* studenti'
	'bresciani'	'bresciane/i'	'bresciane/i/3' or 'bresciane/i/u'	'del bresciano', 'di area bresciana', or 'di Brescia'	–

Table 3

Example of portions of texts with cases of overextended gender (*span*) annotated by each annotator (*ann.*).

alternatives.

To summarize, the annotation scheme has proven valuable both practically and theoretically. It facilitated the identification of gender-unfair expressions and the formulation of alternatives. Moreover, it revealed the inadequacy of an exclusive distinction between obscuration and visibility, emphasizing the need to incorporate a new type of strategy (i.e., 'mixed' alternatives) into the classification.

Although the created annotation scheme has been applied only to administrative texts so far, the guidelines are formulated in such a way that they can be easily applied to data pertaining to different domains. Indeed, we plan to extend the annotation to other data, like web pages of a University that describes its organization and its events. Finally, the resulting corpus, composed of 3,195 portions of texts identified as gender-unfair and reformulated with at least one alternative, could be used in the context of training models to recognize gender-unfair expressions and suggest their alternatives.

Ethics Statement

The annotators have been paid in the context of the actions provided by the Gender Equality Plan of the University of Brescia. The time of annotation has been monitored to ensure that the actual time spent annotating matched the agreed-upon paid hours.

Limitations

Our work presents some limitations. Firstly, the sample of analyzed texts is small and related to a specific domain. To test the robustness of the proposed guidelines, we planned to expand this corpus and its analysis. Secondly, in this work we presented an annotation schema to recognize gender-unfair language and to reformulate it, specifically for Italian, limiting its adaptation to other languages.

Acknowledgments

The annotation has been funded by Università degli Studi di Brescia as part of the actions provided for by the Gender Equality Plan.

The authors would like to extend special thanks to Alessandra Luccioli and Manuela Sanguinetti for their interest and their precious collaboration in this project.

References

- [1] S. Sczesny, M. Formanowicz, F. Moser, Can gender-fair language reduce gender stereotyping and discrimination?, *Frontiers in psychology* 7 (2016) 25.
- [2] J. Saul, E. Diaz-Leon, S. Hesni, Feminist Philosophy of Language, in: E. N. Zalta, U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, Fall 2022 ed., Metaphysics Research Lab, Stanford University, 2022.
- [3] P. Gygax, U. Gabriel, O. Sarrasin, J. Oakhill, A. Garnham, Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men, *Language and cognitive processes* 23 (2008) 464–485.
- [4] P. Gygax, S. Sato, A. Öttl, U. Gabriel, The masculine form in grammatically gendered languages and its multiple interpretations: a challenge for our cognitive system, *Language Sciences* 83 (2021) 101328.
- [5] MIUR, Linee guida per l'uso del genere nel linguaggio amministrativo del MIUR, MIUR, 2018. URL: https://www.miur.gov.it/documents/20182/0/Linee_Guida_per_l_uso_del_genere_nel_linguaggio_amministrativo_del_MIUR_2018.pdf/3c8dfbef-4dfd-475a-8a29-5adc0d7376d8?version=1.0&t=1520428640228.
- [6] G. Attanasio, S. Greco, M. L. Quatra, L. Cagliero, R. Raus, M. Tonti, T. Cerquitelli, E-mimic: Empowering multilingual inclusive communication, *IEEE Big Data Workshops* (2021).
- [7] G. Sulis, V. Gheno, The debate on language and gender in Italy, from the visibility of women to inclusive language (1980s-2020s), *The Italianist* 42 (2022) 153–183. doi:10.1080/02614340.2022.2125707.
- [8] A. Sabatini, *Il sessismo nella lingua italiana. raccomandazioni per un uso non sessista della lingua italiana*, Rome: Istituto Poligrafico e Zecca dello Stato (1987).
- [9] C. Robustelli, Donne grammatica e media. con prefazione di n. maraschio, presidente accademia della crusca, in: *Donne Grammatica e Media*, ITA, 2014, pp. 1–79.
- [10] A. M. Thornton, Per un uso della lingua italiana rispettoso dei generi, Università degli Studi dell'Aquila, 2020. URL: <https://www.univaq.it/include/utilities/blob.php?item=file&table=allegato&id=4925>.
- [11] V. Gheno, Lo schwa tra fantasia e norma, *La Falla del Cassero* (2020). URL: <https://lafalla.cassero.it/lo-schwa-tra-fantasia-e-norma/>.
- [12] I. González Vázquez, A. Klieber, M. Rosola, Beyond pronouns. gender visibility and neutrality across languages, in: L. Anderson, E. Lepore (Eds.), *The Oxford Handbook of Applied Philosophy of Language*, Oxford University Press, forthcoming.
- [13] G. Marcato, E.-M. Thüne, Gender and female visibility in Italian, in: M. Hellinger, H. Bußmann (Eds.), *Gender Across Languages*, Vol.2, John Benjamins Publishing, 2002, pp. 187–217.
- [14] F. Formato, *Gender, Discourse and Ideology in Italian*, Palgrave Macmillan, 2019.
- [15] M. Dardano, P. Trifone, *La nuova grammatica della lingua italiana*, Zanichelli, 1985.
- [16] M. R. Costa-jussà, An analysis of gender bias studies in natural language processing, *Nature Machine Intelligence* 1 (2019) 495–496.
- [17] A. A. Rescigno, E. Vanmassenhove, J. Monti, A. Way, A case study of natural gender phenomena in translation a comparison of google translate, bing microsoft translator and deepl for english to Italian, french and spanish, *Computational Linguistics CLiC-it 2020* 359 (2020).
- [18] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, J. Zhao, D. Mirza, E. Belding, K.-W. Chang, W. Y. Wang, Mitigating gender bias in natural language processing: Literature review, *arXiv preprint arXiv:1906.08976* (2019).
- [19] K. Stanczak, I. Augenstein, A survey on gender bias in natural language processing, *arXiv preprint arXiv:2112.14168* (2021).

Appendix

Type of term	Number	Innovative										Double masculine and feminine form separated by a period	Plural in "-ai"
		Star	u	Schwa	Suffix omission	Underscore	Dash	Apostrophe	At sign	x	y		
Nouns whose gender is overtly signaled by an inflectional suffix	sg			Maestro	Maestr	Maestr'	Maestr@	Maestrx	Maestry		Maestrao; Maestroa Maestrei; Maestrie	-	
	pl	Maestr*	Maestru	Maestr3	Maestr	Maestr_	Maestr@	Maestrx	Maestry		Maestrao; Maestroa Maestrei; Maestrie	Maestrai	
Nouns whose gender is overtly signaled by a derivational suffix	sg	Letto*; Lettric*	Lettonu; Lettricu	Letto3; Lettric3	Letto; Lettric	Letto_; Lettric_	Letto@; Lettric@	Lettox; Lettrix	Lettoy; Lettricy		Letto3; Lettric3; Letto3; Lettric3	-	
	pl	Letto*; Lettric*	Lettonu; Lettricu	Letto3; Lettric3	Letto; Lettric	Letto_; Lettric_	Letto@; Lettric@	Lettox; Lettrix	Lettoy; Lettricy		Letto3; Lettric3; Letto3; Lettric3	Lettorai; Lettrici	
Nouns whose feminine gender is overtly signaled by a derivational suffix	sg	Dotto* Dotto3	Dotto	Dotto3	Dotto	Dotto_	Dotto@	Dotto	Dotto		Dotto3; Dotto3	-	
	pl	Dotto* Dotto3	Dotto	Dotto3	Dotto	Dotto_	Dotto@	Dotto	Dotto		Dotto3; Dotto3	Dottorai	
Nouns whose masculine gender is signaled by opposition to the feminine correspondent	sg	Ero3	Ero3	Ero3	Ero	Ero_	Ero@	Erox	Eroy		Ero3; Ero3	-	
	pl	Ero*	Erou	Ero3	Ero	Ero_	Ero@	Erox	Eroy		Ero3; Ero3	Eroi	
Semi-epicene nouns, whose gender is overtly signaled in the plural but not in the singular	sg	Atlet*	Atletu	Atlet3	Atlet	Atlet_	Atlet@	Atletx	Atlety		Atleti; Atletie	Atletai	
	pl	Atlet*	Atletu	Atlet3	Atlet	Atlet_	Atlet@	Atletx	Atlety		Atleti; Atletie	Atletai	
Epicene nouns, whose gender is not overtly signaled	sg						Docente	Docenti					
	pl						Docenti	Docenti					
Definite articles	sg	L*	Lu	Lo	L?	L_	L@	Lx	Ly		ila?; Loa?; La.o? Ghile?; ile?; Lei?	-	
	pl	L*; L*;	Lur;u	L3;3	L?; L?;	L_? L_?;	L@; L@;	Lx;x	Ly;y		ila?; Loa?; La.o? Ghile?; ile?; Lei?	Lai	
Indefinite articles	sg	Un*	Unu, un*	Uno, un*	Un, un	Un_	Un@, un*	Unx, un*	Uny, un*		Un.a, un.?	-	
	pl	Un*	Unu, un*	Uno, un*	Un, un	Un_	Un@, un*	Unx, un*	Uny, un*		Un.a, un.?	-	
Unverbedated preposition + definite article	sg	Dell*	Dellu	Dello	Dell?	Dell_	Dell@	Dellx	Delly		Della.o; Dello.a; Della	-	
	pl	Dell*; De*	Dellu; Deu	Dell3; Deo	Dell?; De?	Dell_; De_	Dell@; De@	Dellx; Dex	Delly; Dey		Della.o; Dello.a; Della	Dellai	
3.sg pronouns	sg	L'i	?	Lai	?	L_i	L@i	Lxi	Ly?i		Leui?; Lui?; Luet?; Lui.ai?	Lai	
	pl	L'i	?	Lai	?	L_i	L@i	Lxi	Ly?i		Leui?; Lui?; Luet?; Lui.ai?	Lai	

Table 4

Comprehensive table with the alternatives for innovative forms in Italian. We underline that these alternatives can also be used for adjectives or forms of past participle with a gender-marked suffix. Moreover, with the "?" symbol, we refer to the alternatives whose actual form is not totally clear nor, until now, well identified by Italian speakers.

Testo in esame:

- Sensibilizzare i **Rappresentanti degli Studenti** a interloquire direttamente con i **Docenti** laddove ci siano delle necessità specifiche che possono essere facilmente risolte (es. dimensione carattere slides).

Quale tipologia di discrepanza presentano le parole del testo?
 studenti, rappresentanti, docente

genere incongruo
 overesteso
 generico
 non so

Visibilità Conservativa

Visibilità Innovativa

Discouramento Conservativo

Discouramento Innovativo

Alternative ibride

Altre parole, oltre a quelle di riferimento, che presentano discrepanza

Commenti

Figure 1: Label Studio set up with a text containing an overextended and a generic gender-unfair text span.

Testo in esame:

La Direttrice del dipartimento è la prof.ssa ~~illegibile~~, **ordinario** per il Settore scientifico Disciplinare IUS/08 - Diritto Costituzionale.

Quale tipologia di discrepanza presentano le parole del testo?
prof.ssa, direttrice, ordinario

genere incongruo 1 | **sovraesteso** 2 | **generico** 3 | **non so** 4

Alternativa Congrua

Altre parole, oltre a quelle di riferimento, che presentano discrepanza

Commenti

Figure 2: Label Studio set up with a text containing an incongruous gender-unfair text span.