University Speaking for Everyone: Assessing Changes in Italian Higher Education Statutes Toward Gender-Inclusive Language

Sebastiano Vecellio Salto[☼☾] **Camilla Casula**[☼]
Alessio Palmero Aprosio[☾] **Sara Tonelli**[☼]
{svecelliosalto, ccasula, aprosio, satonelli}@fbk.eu
[☼] Fondazione Bruno Kessler, Italy
[☾] University of Trento, Italy

Abstract

We examine the editorial evolution of Italian university statutes toward inclusive language, analyzing how institutions represent female and non-binary identities and how these representations affect administrative communication. To this end, we compile and annotate a corpus of university statutes, tracing the changes that have led some universities to move from the use of the generic masculine to more inclusive formulations. We also experiment with tools for the automatic detection of non-inclusive language in institutional communication and methods for the automatic rewriting of texts into inclusive language.

Keywords: inclusive language, text rewriting, administrative communication

1. Introduction

Language is crucial for conveying information about individual characteristics and group membership in society, with gender being frequently used either as a tool for inclusion or as a means of reinforcing social disparities (Stahlberg et al., 2007; Hovy and Spruit, 2016). Inclusive language policies have therefore become an important concern for public institutions aiming at ensuring equality and representation in their communication.

In the context of Italian, gender-inclusive language still poses a challenge in a number of applications. For instance, a long-standing habit of addressing only males in administrative documents, following the *male as norm* bias, has traditionally been justified with the absence of grammatically gender-neutral forms, as well as the supposed convenience and 'inclusiveness' of a "generic masculine" form, referred to as "unmarked neutral masculine" (Sabatini and Mariani, 1987) or as "overextended masculine" (Cerabolini et al., 2024).

With the publication of official national gender-equality recommendations by Sabatini and Mariani (1987), the Italian public administration (PA) began a conscious shift toward the use of more inclusive language, intended to signal membership to those not represented by a masculine-centered language. Universities, along with municipal and provincial administrations, often independently propose regulations drafted by experts and tailored to their specific needs, see for example Comune di Trento (2021). These guidelines, however, are non-prescriptive and often overlooked, even within the very institutions that propose them, making it difficult to assess their adoption and actual impact.

University statutes provide an interesting case study to investigate these aspects. Universities are expected to promote gender equality, often through dedicated committees and policies, as they serve as both training grounds for future public employees and centers of linguistic and social research. However, these same institutions often perpetuate biases in their communication. The goal of this work is therefore to analyse the inclusiveness of the language used in Italian university statutes, where institutional roles, governance structures and core values are defined. As statutes share comparable structures and content, making them suitable for cross-university analyses, we provide an in-depth statistical examination of the different strategies adopted in their drafting. Given the ongoing debate on the most appropriate and effective forms of gender-inclusive Italian (Fiorentini and Oggionni, 2024), this case study contributes to monitoring the recent impact of the Italian PA's efforts toward gender equity.

In our study we investigate three main research questions: *1)* To what extent are Italian university statutes gender-inclusive? *2)* What gender-inclusive strategies do they adopt? *3)* Can we automatically detect non gender-inclusive language in statutes and rewrite them effectively with LLMs?

Our analysis is not exclusively focused on gender-neutral language – avoiding gendered expressions altogether in Italian (Piergentili et al., 2025) or employing neo-morphemes (Comandini, 2024) – as such strategies are not widely adopted in university statutes and are, in some cases, even discouraged (ITTIG – CNR e Accademia della Crusca, 2011). We aim to account for both gender-inclusive forms, which explicitly represent feminine and masculine identities, and gender-neutral expressions.

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 4058–4069
11–16 May 2026.

2. Background

Inclusive Language Guidelines Today, most entities within the PA have adopted linguistic guidelines for official documents, e.g., European Parliament (2018) and MIUR (2018).[1] These documents offer recommendations and practical examples to promote clear and inclusive communication, while increasing the representation of feminine identities. A common recommendation is to make female identities visible by avoiding using only 'inclusive' masculine nouns, instead pairing them with feminine words (e.g., "*il e la Presidente*", "the$_{m\&f}$ President") and declining official positions,[2] while avoiding unnecessary feminine markedness (e.g., "*la vigile*", not "*la vigilessa*" or "*la vigile donna*", "the policewoman"). This can be done by expressing both genders, either by jointly using masculine and feminine forms (e.g., "*bambini e bambine*", "children$_{m\&f}$"), or by merging roots and appending the corresponding feminine or masculine affixes separated by a slash (e.g., "*il/la bambino/a*", "the$_{m/f}$ child$_{m/f}$"). However, such forms are often considered cumbersome (Robustelli, 2023), and a common recommendation across all guidelines is to resort to collective nouns that imply neutrality and inclusivity, or syntactic structures that avoid gender-marking elements, thereby obscuring gender information.

Italian University Statutes In 2024, the University of Trento published a statute drafted using its own inclusive lexicon, accompanied by regulations written with generic feminine official titles (Università degli Studi di Trento, 2017, 2024,a). Presented as an effort to revive the discussion on inclusiveness (Università degli Studi di Trento, 2024b), the initiative was controversial (Castigliani, 2024), but drew attention to the implications of statutes for representation, as legal documents that address employees and students. Although 22 of Italy's nearly one hundred universities are headed by female rectors, statutes customarily refer to the office using the masculine "*il Rettore*," "the$_m$ Rector$_m$." While respecting the gender of officeholders may seem formal, ensuring inclusive language is particularly important for students (Pepponi, 2023), as generic masculine forms have been shown to cause feelings of exclusion for many identities (Zanoli, 2022).

[1] All referenced guidelines are listed in bibliography.

[2] In Italy, the use of masculine and feminine forms for prestigious positions is debated (Zanoli, 2022; Zarra, 2023). Some favor feminine forms to highlight women's achievements, while others stick to traditional masculine ones to imply parity with male colleagues, or resist new formulations. The debate is particularly active in politics, where self-identification carries ideological weight, as exemplified by Giorgia Meloni referring to herself as "*il presidente del Consiglio*", "the$_m$ Prime Minister$_m$."

Natural Language Processing Automation of generic masculine detection and gender-inclusive rewriting of lengthy administrative documents could make adherence to official guidelines easier (Pepponi, 2023). However, most solutions focus on English, while languages requiring gender agreement such as Italian are less supported (Cerabolini et al., 2024).

As regards machine translation, efforts have been focused on gender-neutral rewriting. While eliminating gender signals supports non-binary inclusion, such strategies often conflict with institutional guidelines, which discourage radical changes that may confuse users (Iannizzotto and Di Valvasone, 2023). However, although there are trade-offs between neutrality and meaning preservation, these tools can be effective (Piergentili et al., 2025).

The E-MIMIC project (Attanasio et al., 2021) has developed `Inclusively`, a tool for identifying and rewriting non-inclusive language in Italian, thanks to a corpus of annotated administrative documents (Greco et al., 2025). NLP resources for Italian universities have also been proposed by Cerabolini et al. (2024), who draft a dictionary of biased words, a database of administrative sentences rewritten inclusively, a rule-based POS-tag-based approach, a fine-tuned `mT5` model, and a curated prompt for `GPT-4`. Riccio (2024) analyzed the frequency of titles and institutional roles within the university regulatory corpus (in Italian, *Corpus Normativo Universitario*), a collection of statutes, regulations, and other documents from 91 Italian universities.

In this work, we focus on statutes at the level of individual institutions, comparing each statute's inclusive and non-inclusive versions to identify preferred gender-inclusion strategies and assess whether these can be systematically recognized and applied. Furthermore, all experiments are based on real documents from Italian institutions annotated using a curated taxonomy, allowing us to approach multiple kinds of gender-inclusive rewriting.

3. Corpus

We present a corpus of statutes of Italian higher education institutions and, along with it, we curate a subcorpus containing all documents available in both non-inclusive and inclusive versions. Each variation between the two versions is annotated for non-gender related changes and various kinds of gender representation, so to enable an in-depth analysis of how Italian universities implement inclusive language.

3.1. Data Collection

We first collected the available statutes of all Italian higher education institutions from their official web-

University	N.I.	Inc.
Scuola Normale Superiore (*sns*)	2019	2022
Uni. degli Studi di Brescia (*unibs*)	2020	2024
Libera Università di Bolzano (*unibz*)	2002	2003
Uni. per Stranieri di Siena (*unistrasi*)	2021	2024
Uni. degli Studi di Trento (*unitn*)	2012	2024
Uni. degli Studi della Tuscia (*unitus*)	2019	2024
Uni. Ca'Foscari Venezia (*unive*)	2013	2023

Table 1: List of universities in our subcorpus, with the publication year of the last non-inclusive statute and of the first inclusive statute.

sites, obtaining 99 documents. We then manually assessed whether they adopted gender-inclusive language, finding that only 11 employed 'some degree of inclusivity', defined as language extending beyond the sole use of masculine forms (e.g., by including both feminine and masculine terms).

For each statute in this subset, we retrieved the most recent version written exclusively in the masculine form, creating a parallel corpus with two versions per document: masculine-only vs. (more) inclusive. We had to discard three pairs of statutes from the parallel corpus – from the universities of Bergamo (2012; 2023), Siena (2005; 2024), and Scuola Superiore Sant'Anna (2018; 2024) – because, despite being inclusive, they could not be adequately aligned with the previous version due to extensive rewriting and restructuring. We also excluded the statutes of university of Palermo (2019; 2022), as they contained only minimal inclusive language. The final dataset thus includes both versions of the statutes of 7 universities (see Table 1), containing nearly 200,000 tokens in total.

We observe that the first gender-inclusive version of the bilingual statute of university of Bolzano dates back to 2003, nearly twenty years ahead of other institutions. As noted by Chiocchetti and Ralli (2023), the Autonomous Province of Bolzano, whose administrative communication is officially conducted in both Italian and German, has been influenced by how German-speaking countries employ non-discriminatory writing strategies.[3]

3.2. Data Annotation

Since our goal is to compare original and rewritten texts to analyse the different linguistic choices and rewriting strategies adopted by each university, we automatically extract changes between the non-inclusive and the inclusive version of each statute. To do this, we merge the two versions into a single JSON file using the `diff` Linux package, which

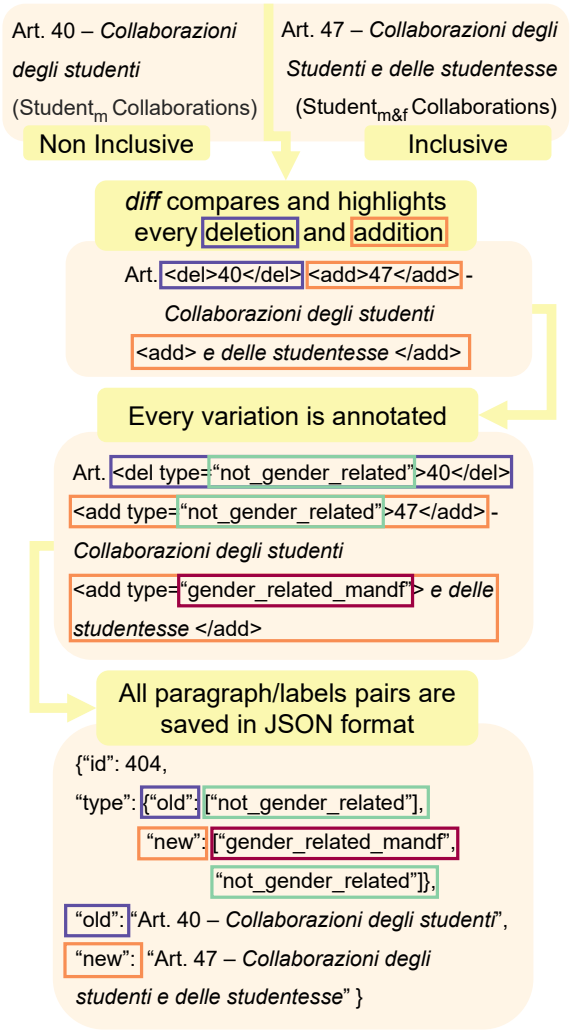

Figure 1: Detection and labelling of differences between versions of a university statute, comparing the same article from *unitn* (2012 and 2024).

compares two texts at paragraph level. The package signals each difference between the two versions, marking text spans that were added to the inclusive document (`<add></add>`) or deleted from its non-inclusive version (`<del></del>`).

This allows us to focus annotation efforts only on text spans that changed, rather than annotating the entire statutes, which would be extremely time-consuming. While the use of `diff` introduced some minor noise in cases where substantial structural revisions hindered perfect paragraph alignment, we consider such instances to have a negligible effect on the validity of our analysis.

Each pair of non-inclusive/inclusive paragraphs is then labeled to distinguish between changes that were not related to gender representation from those concerning inclusive language, as summarized in Figure 1. We first automatically label each pair of aligned paragraphs as **no_change** when they are identical. Then, we ask an annotator

[3]We suppose that bilingualism also influences the strategies adopted, as German readily accommodates forms that are equally suitable for Italian.

to manually label the remaining pairs to account for the different types of changes. We use the **not_gender_related** label for changes relating to lexicon or syntax, punctuation, article numbering, or changes in university governance. A second category, **gender_related**, comprises changes that introduce, remove, or alter gender representation, and is further specified as: **m**, when the annotated text includes only masculine forms (in deletions, this refers to the removal of generic masculine forms; in additions, to their non-inclusive introduction); **f** for generic feminine forms with no masculine representation; **fslashm** or **mslashf** for feminine/masculine or masculine/feminine forms separated by a slash (e.g., "*lo/a studente/ssa*", "the student$_{m/f}$"); **fandm** or **mandf** for joint feminine and masculine terms (e.g., "*gli studenti e le studentesse*," "the students$_{m\&f}$"); **group** for gender-neutral collective nouns (e.g., "*la classe studentesca*," "the student body"); and **obscuring** for cases where gender is no longer explicitly expressed (e.g., "*chi frequenta l'università*," "who attends the university," instead of "*gli studenti*," "the students$_{m}$"). A single text span can be assigned multiple labels. As this task is inherently objective, we consider our annotation reliable and do not involve a second annotator, following the approach adopted in Attanasio et al. (2021) on a similar task.

3.3. Corpus Statistics

We release a corpus comprising the statutes of all Italian universities in PDF format, along with annotated JSON files for the seven universities we analyzed.[4] Since today only 11% of the statutes considered are at least partially gender-inclusive and, except for *unibz*, they were published from 2022 onward, our corpus aims to serve as a resource for monitoring this ongoing transition towards inclusive language.[5] Below, we provide statistics about the content of the annotated subcorpus.

Gender-related vs. non-gender-related changes For each university, we report in Table 2 the total number of paragraphs (¶) in the statute. Most paragraphs have undergone changes, primarily of structural nature, annotated as **not_gender_related** (**NGR**). This suggests that the primary motivation behind the updates was not representation, but administrative considerations. Given the routine need to revise such documents, we argue that opportunities to improve representation arise naturally and require few additional resources, since gender is relevant (**GR**) in a limited number of paragraphs.

University	¶	NC	NGR	GR
sns	533	36%	44%	38%
unibs	373	19%	74%	39%
unibz	306	17%	75%	29%
unistrasi	959	23%	63%	42%
unitn	442	27%	51%	45%
unitus	528	19%	76%	32%
unive	567	44%	48%	22%
total	**3,708**	**27%**	**60%**	**36%**

Table 2: Distribution of **no_change** (**NC**), **not_gender_related** (**NGR**), and **gender_related** (**GR**) labels at the paragraph level. Since a paragraph may include multiple types of variation, the totals for each university may exceed 100%.

Gender-related paragraphs For each paragraph containing gender-related changes, we report the distribution of gendered forms in the inclusive version of each statute in Table 3. No instances of generic **f**eminine expressions are found across any of the statutes, while a few remain exclusively **m**asculine, often in statutes that retain masculine forms for official titles such as "*Rettore*," "Rector$_{m}$." Although often accompanied by clauses specifying that such titles apply to both masculine and feminine identities, we consider this a sub-optimal inclusive choice. This suggests that feminine visibility results from deliberate efforts toward representation, never from the elimination of masculine counterparts. By contrast, habitual usage and limited familiarity with inclusive language appear to sustain the persistence of generic masculine forms.

We observe that expressing both masculine and feminine forms separated by a slash is slightly more frequent than joint expression. This contrasts with most institutional guidelines, which recommend the former mainly for shorter texts such as forms (*i.a.*, MIUR (2018)). All universities, except for *sns*, adopt a "masculine/feminine" order. While this may sometimes be morphologically convenient, since many feminine words are derived from masculine roots (e.g., "*lo/a studente/ssa*," "the student$_{m/f}$"), alternating this habit could help avoid reinforcing masculine preemption. When both genders are jointly expressed, the preferred order is also "masculine and feminine," though this varies across universities. Finally, we note the limited use of collective nouns and gender-obscuring strategies, representing a missed opportunity to ease the representation of non-binary identities.

Each university tends to favor one inclusion strategy, yet all employ multiple ones. We consider the

[4] https://github.com/dhfbk/statutes

[5] This shift appears to be primarily driven by institutional guidelines, whose proposed strategies can be clearly recognized. By contrast, other factors appear to contribute very little, as even the presence of female rectors does not seem to correlate with more inclusive official communication.

Univ.		And		Slash		Neutral	
	m	**m&f**	**f&m**	**m/f**	**f/m**	**gr.**	**ob.**
sns	1%	0%	14%	0%	**25%**	1%	1%
unibs	1%	23%	4%	0%	0%	**12%**	**6%**
unibz	3%	0%	0%	**25%**	0%	1%	1%
unistrasi	0%	10%	0%	**25%**	0%	1%	0%
unitn	1%	13%	8%	**20%**	2%	9%	2%
unitus	6%	0%	11%	0%	0%	**16%**	2%
unive	4%	2%	0%	**2%**	0%	8%	1%
average	**2%**	**7%**	**5%**	**10%**	**4%**	**7%**	**2%**

Table 3: Percentage of paragraphs in the inclusive version of the statutes that include each gender representation strategy relative to the total number of paragraphs. A single paragraph may contain multiple strategies. **gr.** = group; **ob.** = obscuring.

University	Previous	Current	Delta
sns	14,439	15,388	+7%
unibs	9,219	10,169	+10%
unibz	6,871	7,639	+11%
unistrasi	23,578	25,383	+8%
unitn	11,043	13,580	+23%
unitus	13,189	15,642	+19%
unive	15,078	14,793	**-2%**
total	**93,417**	**102,594**	**+10%**

Table 4: Number of tokens for both statute versions and relative increase.

presence of more than one strategy within a paragraph as an indicator of micro-variability in implementation. Calculated over the number of paragraphs containing at least one **gender_related** label, *unive* has the highest proportion of paragraphs employing a single gender-visibility strategy (98%), followed by *unibz* (93%), *sns* (87%), *unistrasi* (86%), *unibs* (75%), and *unitn* and *unitus* (73%). All inclusive language guidelines adopted seem to provide clear contextual use criteria, while the occasional coexistence of strategies may either reflect uncertainty in their implementation or, conversely, deliberate stylistic variation intended to alternate gender order and reflect openness to linguistic experimentation.

Length In the public debate on gender-inclusive language, a frequent criticism is that it may make texts long and cumbersome to read. To verify this claim, we compare the length of the two versions of each statute. Results are reported in Table 4.

All inclusive statutes are longer than their non-inclusive counterparts, except for *unive*. On average, the increase amounts to approximately 1,300 tokens . Given that statutes are already lengthy documents, this increase should have little impact on their overall readability and accessibility, especially when weighed against the benefit of making them represent *nearly* the entire academic community.

We further analyze the variation in the number of tokens based on the different types of gender-inclusive strategies (see Figure 2). Paragraphs containing gender-related or non-gender-related changes show a comparable percentage increase relative to their length in the older version of the statutes. Within gender-related modifications, we analyze those adhering to a binary representation (in our taxonomy: **m**, **f**, **mandf**, **fandm**, **m/f**, **f/m**) and those that, even within a binary context, are gender-neutral (**group** and **obscuring**). The latter contribute only marginally to the overall lengthening of the documents, as expected given their limited use. Nonetheless, this suggests that greater attention should be devoted to their adoption, rather than dismissing them as space-consuming strategies.

Since *unive* is an outlier, we further examined the information reported in Tables 2 and 4, and Figure 2. The overall reduction in length reflects a deliberate effort toward conciseness, effective in both structural and gender-related revisions. This suggests that inclusive language, although it may entail a relative token increase, as noticeable regarding gender-neutral forms, can be implemented through targeted interventions with minimal impact on the overall length of the text.

4. Experimental setup

As a next step, we investigate whether it is possible to detect gender-inclusive language using NLP and to automatically rewrite non-inclusive text. In particular, starting from the corpus described in the previous sections, we devise an experimental setup in which we use it for three separate tasks:

- **Ternary classification**, in which we aim at detecting whether a text is *inclusive*, *not inclusive*, or it does *not contain references to gender*.
- **Fine-Grained classification**, in which we aim to not only detect whether a text is *inclusive*, but also to identify *which strategies* are used to make the text inclusive, based on the annotation scheme described in Section 3.2.
- **Rewriting**, in which a model takes as input a text and makes it more inclusive if it uses only the masculine form.

For the first two tasks, we test both fine-tuning BERT-based classification models and LLMs in zero- and few-shot scenarios, while for the rewriting task we only use generative LLMs.[6] To increase

[6]While preliminary experiments also included fine-tuning of sequence-to-sequence language models, they showed unsatisfactory results and were thus discarded from our experimental setup.

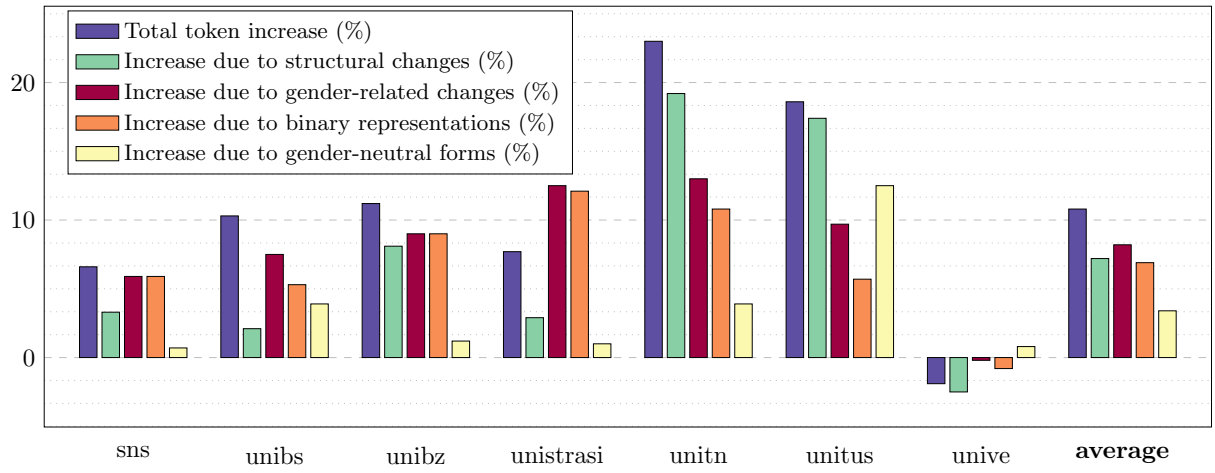

Figure 2: Percentage token increase compared to the previous non-inclusive version by paragraph label type. Binary representations: **m**, **f**, **mandf**, **fandm**, **m/f**, **f/m**. Gender-neutral forms: **group**, **obscuring**.

the size of the training set, we also annotated the statutes from *unipa*. To increase masculine examples, we also annotated the non-inclusive statute from *unisi*, but only for the presence of masculine representations within paragraphs. For all three tasks, we use paragraphs from two universities (*unive* and *unitn*) as the test set, and those from the remaining institutions (*sns*, *unibs*, *unibz*, *unistrasi*, *unitus*, plus *unipa* and *unisi* only for the ternary and fine-grained classification tasks) as the training and development sets, with an 80/20 split.

4.1. Models

For the two classification tasks, we fine-tune three encoder models available on Huggingface: `XLM Roberta Base`,[7] `XLM Roberta Large`,[8] and `BERT Italian xxl`.[9]

For the ternary classification task, the three classifiers are fine-tuned with a LR of 2e-5 and a batch size of 32 for 10 epochs, loading at the end the model from the best epoch in terms of macro-F_1. For the fine-grained classification task, they are fine-tuned with a LR of 1e-5 and a batch size of 16 for 20 epochs, choosing the model from the best epoch in terms of macro-F_1.

For all three tasks, we also test 4 large language models through the OpenAI and Groq APIs: `GPT-4o`,[10] `Llama 3.1 8B Instant`,[11] `Llama 3.3 70B Versatile`,[12] and `Llama 4 Scout 17B 16E Instruct`[13] in both few-shot and zero-shot scenarios. The hyperparameters used for generation are the default ones of the models' respective APIs for all tasks.

In the few-shot settings, we use the training set (see Section 4) as the source of example instances. Specifically, for each sentence in the test set, we retrieved, for every possible output class, the five most similar sentences from the training set, as determined using the `TfidfVectorizer` class provided by `sklearn`.

All prompts and few-shot examples used are available in the GitHub repository (see Section 3.3). Additionally, as a reference baseline, we test the `Inclusively` model both for ternary classification[14] and for rewriting.[15]

4.2. Ternary Classification

We cast the detection of non-inclusive language as a ternary classification task, in which models are expected to predict one of three classes: **not_gender_related**, **m**asculine only, and **inc**lusive. This kind of setup is common in previous work on this topic (see for instance Attanasio et al. (2021)) and allows the quick identification of texts that may benefit from more inclusive wording. To reduce our annotation scheme to three labels only, we aggregate our annotations into these three categories using a set of manually crafted rules.

First, if at least one span in the paragraph contains the annotation **gender_related_m**, we assume that the text is not inclusive, although it might also contain inclusive terminology. As such, we consider the entire paragraph as being tagged with the generic **m**asculine label. If any other **gender_related** annotation is present in the text, it is labeled as **inclusive**, unless the text uses only the

[7] FacebookAI/xlm-roberta-base
[8] FacebookAI/xlm-roberta-large
[9] dbmdz/bert-base-italian-xxl-cased
[10] cdn.openai.com/gpt-4o-system-card.pdf
[11] meta-llama/Llama-3.1-8B
[12] meta-llama/Llama-3.3-70B-Instruct
[13] meta-llama/Llama-4-Scout-17B-16E-Instruct
[14] E-MIMIC/inclusively-classification
[15] E-MIMIC/inclusively-reformulation-it5

obscuring strategy. In fact, in these classification examples, the **gender_related_obscuring** tag is treated as **not_gender_related**, as in themselves these expressions do not refer to gender explicitly. Finally, the **not_gender_related** tag is assigned to a text only if it is the only label for that paragraph.

The final ternary classification dataset includes 3,085 paragraphs in the train and dev sets (of which 52% are tagged as **not_gender_related**, 26% as **inc**, and 22% as **m**). The test set for this task includes 816 paragraphs, of which 49% **not_gender_related**, 29% **inc**, and 22% **m**).

4.3. Fine-Grained Classification

We also experiment with fine-grained classification of non inclusive language, based on seven classes corresponding to the gender-inclusive strategies identified in our taxonomy, namely **not gender related**, **m**, **mslashf**, **fslashm**, **fandm**, **mandf**, and **group**.[16] As our labels are referred to paragraphs, this is a multi-label problems, as multiple strategies might occur within the same paragraph.

Through this fine-grained detection, one could for instance identify texts that are inclusive with regards to masculine and feminine gender representations, but not to non-binary ones, which are frequently represented via the **group** class.

The final fine-grained classification dataset includes 3,064 paragraphs in train and dev and 818 paragraphs in the test set. In the overall dataset, 51% of paragraphs are **not_gender_related**, 34% are **m**, 4% use the **mslashf** strategy, 4% use multiple strategies, and the remaining strategies (**fslashm**, **mandf**, **group**, and **fandm**) are represented in 2% of paragraphs each.

4.4. Rewriting

The third task focuses on rewriting non inclusive texts in a more inclusive way. To create the training data for this task, we first extract from the corpus described in Section 3.3 only the paragraphs for which we have both a non-inclusive and an inclusive version. Then, we further filter the non-inclusive/inclusive text pairs using fuzzy matching with the `thefuzz` library,[17] only preserving pairs whose similarity score is over 0.7, as empirically estimated in a manual analysis.

The final parallel test dataset includes 388 paragraphs, whose inclusive versions employ the **mslashf** and **group** strategies in 25% of cases each, 20% include multiple, 17% **mandf**, 9% **fandm**, 2.5% **obscuring**, and 0.5% **fslashm**. We also release a set of 1,126 additional parallel paragraphs for future research on this topic.

5. Results and Discussion

We report in Table 5 the classification results, and in Table 6 the results for the rewriting task using BLEU and ROUGE-2, as in other works on rewriting (Greco et al., 2025). We compute both since BLEU is a precision-based metric, capturing the proportion of words in the rewriting that are also in the reference text, while ROUGE-2 is recall-oriented, measuring how many bigrams in the reference text are captured in the rewriting.

Concerning classification, `XLM-Roberta large` achieves the best performance both in the ternary and in the fine-grained task, with a Macro-F_1 of .870 in the former and .542 in the latter. Overall, we observe that all the BERT-based fine-tuned classifiers clearly outperform even state-of-the-art LLMs in both zero- and few-shot setups in the two classification tasks. As expected, the fine-grained classification task is more challenging than the ternary one due to the higher number of classes and consequently less examples for each label in the training. Interestingly, many of the errors made by the classifiers in the fine-grained task regard **not_gender_related** and generic **m**asculine texts being confused with each other, perhaps indicating that these classifiers are to some extent subject to the *male as norm* bias. Overall, if we consider both the classification performance and the computational costs, BERT-based classifiers emerge as effective and low-impact tools for this type of task and for future research applications, while generative LLMs seem less effective.

For what concerns the task of rewriting, commercial LLMs can effectively rewrite non-inclusive texts when prompted with a specific inclusion strategy, achieving a ROUGE-2 of .7692, although their performance is not always significantly better than that of `Inclusively`. A manual analysis of the outputs shows that some frequent errors made by LLMs include using unnecessarily marked feminine forms (e.g. LLMs frequently use "**presidentessa*," "*presidentess"), and feminine forms of masculine words not referring to people, being invented by the LLM (e.g. "*i componenti del Collegio/*Collegia*," "the members of the *colleges$_{m/f}$"). Overall, however, this kind of tool shows to be potentially viable as an option to streamline the process of making university statutes more inclusive, especially when paired with tools that can automatically identify portions of text that are not inclusive.

[16] **Obscuring** is not included in this task, similarly to the ternary classification task, as without the non-gender-inclusive version of the paragraph it is not distinguishable from the **not_gender_related** tag, as both contain no gender information in isolation.

[17] https://github.com/seatgeek/thefuzz

Model	Shot		Ternary Classification Task				Fine Grained Classification Task		
			P$^{\pm stdev}$	R$^{\pm stdev}$	F$_1$$^{\pm stdev}$		P$^{\pm stdev}$	R$^{\pm stdev}$	F$_1$$^{\pm stdev}$
Inclusively		**M**	.838$^{.076}$	.840$^{.067}$	.839$^{.074}$	Not applicable to the task.			
		m	.847$^{.070}$	.847$^{.070}$	.847$^{.070}$				
IT BERT		**M**	.868$^{.004}$	**.872**$^{.007}$	**.870**$^{.005}$	**M**	.494$^{.069}$	.354$^{.050}$	.412$^{.029}$
		m	.876$^{.005}$	.876$^{.005}$	.876$^{.005}$	**m**	**.734**$^{.005}$	**.736**$^{.005}$	**.736**$^{.005}$
XLM-R$_{base}$		**M**	.858$^{.007}$	.864$^{.005}$	.860$^{.004}$	**M**	.314$^{.034}$	.312$^{.044}$	.313$^{.030}$
		m	.870$^{.000}$	.870$^{.000}$	.870$^{.000}$	**m**	.722$^{.004}$	.722$^{.004}$	.722$^{.004}$
XLM-R$_{large}$		**M**	**.872**$^{.007}$	.870$^{.006}$	**.870**$^{.004}$	**M**	**.548**$^{.040}$	**.536**$^{.033}$	**.542**$^{.053}$
		m	**.878**$^{.004}$	**.878**$^{.004}$	**.878**$^{.004}$	**m**	**.734**$^{.008}$	**.736**$^{.008}$	**.736**$^{.008}$
Llama3.1	zero	**M**	.505$^{.000}$	.540$^{.000}$	.522$^{.000}$	**M**	.285$^{.000}$	.267$^{.000}$	.276$^{.000}$
		m	.668$^{.000}$	.668$^{.000}$	.668$^{.000}$	**m**	.530$^{.000}$	.509$^{.000}$	.520$^{.000}$
	few	**M**	.695$^{.003}$	.644$^{.002}$	.669$^{.002}$	**M**	.504$^{.001}$	.326$^{.000}$	.396$^{.000}$
		m	.713$^{.002}$	.713$^{.002}$	.713$^{.002}$	**m**	.597$^{.000}$	.558$^{.000}$	.577$^{.000}$
Llama3.3	zero	**M**	.756$^{.063}$	.572$^{.000}$	.651$^{.023}$	**M**	.367$^{.007}$	.413$^{.009}$	.389$^{.008}$
		m	.694$^{.000}$	.694$^{.000}$	.694$^{.000}$	**m**	.542$^{.001}$	.560$^{.001}$	.551$^{.001}$
	few	**M**	.740$^{.115}$	.670$^{.101}$	.703$^{.108}$	**M**	.542$^{.040}$	.448$^{.010}$	.490$^{.022}$
		m	.790$^{.003}$	.790$^{.003}$	.790$^{.003}$	**m**	.628$^{.000}$	.610$^{.001}$	.619$^{.001}$
Llama4	zero	**M**	.777$^{.002}$	.581$^{.002}$	.665$^{.002}$	**M**	.307$^{.012}$	.497$^{.018}$	.379$^{.014}$
		m	.672$^{.002}$	.672$^{.002}$	.672$^{.002}$	**m**	.523$^{.005}$	.561$^{.004}$	.542$^{.004}$
	few	**M**	.662$^{.115}$	.611$^{.106}$	.635$^{.110}$	**M**	.480$^{.001}$	.367$^{.004}$	.416$^{.003}$
		m	.785$^{.002}$	.785$^{.002}$	.785$^{.002}$	**m**	.570$^{.003}$	.559$^{.005}$	.564$^{.004}$
GPT-4o	zero	**M**	.678$^{.001}$	.568$^{.001}$	.618$^{.001}$	**M**	.344$^{.009}$	.482$^{.015}$	.402$^{.012}$
		m	.686$^{.001}$	.686$^{.001}$	.686$^{.001}$	**m**	.506$^{.004}$	.567$^{.005}$	.535$^{.004}$
	few	**M**	.778$^{.001}$	.690$^{.002}$	.732$^{.002}$	**M**	.519$^{.012}$	.452$^{.009}$	.483$^{.009}$
		m	.764$^{.001}$	.764$^{.001}$	.764$^{.001}$	**m**	.623$^{.003}$	.603$^{.002}$	.613$^{.002}$

Table 5: Mean and standard deviation of **M**acro and **m**icro **P**recision, **R**ecall, and **F**$_1$ scores, averaged across five runs for `Inclusively`, `BERT Base Italian XXL Cased`, `XLM-RoBERTa Base` and `Large`, and `GPT-4o`, and three runs for `Llama 3.1 8B Instant`, `Llama 3.3 70B Versatile`, and `Llama 4 Scout 17B 16E Instruct`. LLMs are evaluated using **zero-** and **few-shot** prompting.

Model	BLEU$^{\pm stdev}$	ROUGE-2$^{\pm stdev}$
`Inclusively`	.4926$^{.2733}$	.7098$^{.2347}$
`Llama3.1`	.6092$^{.2581}$	.6570$^{.2320}$
`Llama3.3`	.4146$^{.2114}$	.7380$^{.2019}$
`Llama4`	.5501$^{.2413}$	.6476$^{.2305}$
`GPT-4o`	**.7600**$^{.2038}$	**.7692**$^{.1755}$

Table 6: BLEU and ROUGE-2 scores for reformulation inferences, with standard deviation.

6. Conclusions

In this work, we released: *a)* a corpus of all available, up-to-date statutes from Italian universities, 11 of which are provided in both non-inclusive and inclusive versions; *b)* a subcorpus of 7 university statutes, manually annotated at the paragraph level to capture structural modifications and various types of gender-related changes between non-inclusive and inclusive versions. Our analysis revealed different gender-representation strategies adopted by universities, noting the scarcity of gender-neutral solutions and a moderate increase in text length, challenging common concerns about the impracticality of inclusive language.

We then evaluated various classifiers and LLMs on three tasks: automatically identifying non-inclusive text extracts, automatically detecting the inclusiveness strategies employed to convey gender information, and generating inclusive rewrites, with the goal of facilitating the transition towards

inclusive language in Italian university statutes. We found that the best approaches for detection tasks consists in fine-tuning of pre-trained encoder models, while various commercial LLMs perform acceptably at rewriting non-inclusive texts when given a specific gender inclusion strategy.

7. Limitations

Our study is exploratory in nature and its scope is constrained by the small number of inclusive statutes available in the main corpus and by the inclusion of only seven annotated university statutes in the subcorpus. Additionally, while the use of the `diff` package enables rapid and automated comparisons, it may have introduced minor noise in cases where substantial structural revisions prevented precise paragraph alignment. Annotation was conducted by a single annotator. Although the task is largely objective in nature, this may have introduced unverified human errors.

8. Acknowledgments

This work has been partly supported by the project "AIxPA – Artificial Intelligence for Public Administration", funded within the Flagship Project PNC-A.1.3 – Digitalization of the Public Administration of the Autonomous Province of Trento, CUP: C49G2200102000.

Bibliographical References

Giuseppe Attanasio, Salvatore Greco, Moreno La Quatra, Luca Cagliero, Michela Tonti, Tania Cerquitelli, and Rachele Raus. 2021. E-MIMIC: Empowering Multilingual Inclusive Communication. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 4227–4234, Orlando, FL, USA. IEEE.

Autonome Provinz Bozen - Südtirol | Provincia autonoma di Bolzano - Alto Adige. 2024. Direttive per il linguaggio di genere. https://assets-eu-01.kc-usercontent.com/e6f8f5ff-b94b-01fb-b36f-926b5e8f8000/123261ed-e32f-4934-beea-67f6a8db3bb4/Richtlinien_ita_nov2024_DEF.pdf. Accessed: 30-09-2025.

Christian Baraldi, Renato Busarello, Olivia Pinto, Alessandra Cervellati, Alessia Graziati, Gaia Minella, Erika Vassallo, and Pierpaolo De Gennaro. 2023. Parole che fanno la differenza. Scrivere e comunicare rispettando le differenze di genere. Comune di Bologna.

Gabriella Bianciardi and Marilena Parisi. 2017. Linee Guida per un utilizzo non discriminatorio del linguaggio in base al genere nell'attività amministrativa. http://www.comune.torino.it/ediliziaprivata/normativa/pdf/all1-DelGC201703517.pdf. Città di Torino. Accessed: 30-09-2025.

Martina Castigliani. 2024. Università di Trento: stop all'uso del maschile anche per le donne, l'ateneo sceglie "il femminile sovraesteso". *Il Fatto Quotidiano*. Visited on 30/09/2025.

Aurora Cerabolini, Gabriella Pasi, and Marco Viviani. 2024. Automating gender-inclusive language modification in italian university administrative documents. In *Natural Language Processing and Information Systems*, pages 333–347, Cham. Springer Nature Switzerland.

Elena Chiocchetti and Natascia Ralli. 2023. Comunicazione istituzionale inclusiva in alto adige: esperienze e questioni aperte. In *CILGI5 2023 Convegno Internazionale di Linguistica e Glottodidattica Italiana "Comunicare l'Italia mediante l'Italiano"*, Salerno.

Gloria Comandini. 2024. Salve a tuttə, tutt*, tuttu, tuttx e tutt@: l'uso delle strategie di neutralizzazione di genere nella comunità queer online. Indagine su un corpus di italiano scritto informale sul web. *Testo e Senso*, pages 43–64.

Comune di Trento. 2021. Linee guida per l'utilizzo del genere nel linguaggio amministrativo. https://trentogiovani.it/content/download/19548/653360/file/linee%20guida%20linguaggio%20di%20genere.pdf. Accessed: 30-09-2025.

Sara De Vido and Gianluca Briguglia. 2024. Linee guida di Ateneo 2024 per un linguaggio non discriminatorio e rispettoso del genere. Università Ca'Foscari Venezia.

European Parliament. 2018. La neutralità di genere nel linguaggio usato al Parlamento europeo. https://m4.ti.ch/fileadmin/CAN/SGCDS/pari_opportunita/download/GNL_Guidelines_IT-original.pdf. Accessed: 30-09-2025.

Ilaria Fiorentini and Rachele Oggionni. 2024. 'Anche la lingua deve adeguarsi.' La percezione pubblica del dibattito sul linguaggio inclusivo. In Anna-Maria De Cesare and Giuliana Giusti, editors, *Lingua inclusiva: forme, funzioni, atteggiamenti e percezioni*, pages 179–202. Fondazione Università Ca' Foscari, Venice.

Freie Universität Bozen | Libera Università di Bolzano | Università Liedie de Bulsan. 2002.

STATUTO della Libera Università di Bolzano | STATUT der Freien Universität Bozen. Accessed: 30-09-2025.

Freie Universität Bozen | Libera Università di Bolzano | Università Liedie de Bulsan. 2003. STATUTO della Libera Università di Bolzano | STATUT der Freien Universität Bozen. Accessed: 30-09-2025.

Silvia Giorcelli, Maria Spanò, Rachele Raus, Miriam Abouyaala, Igor Catrano, and Viviana Patti. 2015. Un approccio di genere al linguaggio amministrativo. https://www.unistrasi.it/public/articoli/318/Files/DR%20396.%2020%20Linee%20guida%20linguaggio%20rispettoso%20del%20genere.pdf. Università degli Studi di Torino.

Salvatore Greco, Moreno La Quatra, Luca Cagliero, and Tania Cerquitelli. 2025. Towards AI-Assisted Inclusive Language Writing in Italian Formal Communications. *ACM Transactions on Intelligent Systems and Technology*, 16(4):1–24.

Dirk Hovy and Shannon L. Spruit. 2016. The Social Impact of Natural Language Processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 591–598, Berlin, Germany. Association for Computational Linguistics.

Stefania Iannizzotto and Luisa Di Valvasone. 2023. La comunicazione istituzionale dell'università tra semplificazione linguistica e strategie possibili per un linguaggio inclusivo. In Maria Paola Monaco, editor, *La lingua italiana in una prospettiva di genere: atti del Seminario online promosso dagli Atenei di Firenze e Udine con il patrocinio dell'Accademia della Crusca (1° marzo 2022)*, pages 69–80. Firenze University Press, Firenze.

ITTIG – CNR e Accademia della Crusca. 2011. *Guida alla redazione degli atti amministrativi. Regole e suggerimenti*. Ittig – CNR, Firenze.

MIUR. 2018. Linee Guida per l'uso del genere nel linguaggio amministrativo del MIUR. https://www.mim.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. Accessed: 30-09-2025.

Donatella Mostacchi and Giovanna Pensabene. 2023. Linee guida per un linguaggio amministrativo rispettoso del genere nella Città metropolitana di Milano. https://www.cittametropolitana.mi.it/export/sites/default/newsletter/doc/Opuscolo-linguaggio.pdf. Città metropolitana di Milano. Accessed: 30-09-2025.

Elena Pepponi. 2023. Lingua... di che genere? riflessioni preliminari e proposte operative per una comunicazione istituzionale universitaria che superi il binarismo di genere. In Maria Paola Monaco, editor, *La lingua italiana in una prospettiva di genere: atti del Seminario online promosso dagli Atenei di Firenze e Udine con il patrocinio dell'Accademia della Crusca (1° marzo 2022)*, pages 81–94. Firenze University Press, Firenze.

Andrea Piergentili, Beatrice Savoldi, Matteo Negri, and Luisa Bentivogli. 2025. Gender-Neutral Rewriting in Italian: Models, Approaches, and Trade-offs. *identities*, 29:4.

Anna Riccio. 2024. Strategie linguistiche per la valorizzazione delle differenze di genere nei testi normativi delle università italiane. Un'analisi corpus-based. In Anna-Maria De Cesare and Giuliana Giusti, editors, *Lingua inclusiva: forme, funzioni, atteggiamenti e percezioni*, pages 19–73. Fondazione Università Ca' Foscari, Venice.

Cecilia Robustelli. 2012. Linee guida per l'uso del genere nel linguaggio amministrativo. https://www.mim.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. Accessed: 30-09-2025.

Cecilia Robustelli. 2023. La lingua italiana come strumento di parità di genere nelle università. In Maria Paola Monaco, editor, *La lingua italiana in una prospettiva di genere: atti del Seminario online promosso dagli Atenei di Firenze e Udine con il patrocinio dell'Accademia della Crusca (1° marzo 2022)*, pages 43–55. Firenze University Press, Firenze.

Martina Rosola, Marta Cremaschi, Camilla Federici, and Alessandra Viviana. 2025. Vademecum sul linguaggio di genere. https://www.unibs.it/sites/default/files/2023-06/Vademecum%20Linguaggio%20Inclusivo_DEF.pdf. Università degli Studi di Brescia. Accessed: 30-09-2025.

Alma Sabatini and Marcella Mariani. 1987. *Il sessismo nella lingua italiana*. Presidenza del Consiglio dei Ministri, Roma.

Scuola Normale Superiore. 2019. Statuto della Scuola Normale Superiore. https://www.sn

s.it/sites/default/files/2020-11/statutoscuolanormalesuperiore.pdf. Accessed: 30-09-2025.

Scuola Normale Superiore. 2022. Statuto della Scuola Normale Superiore. https://www.sns.it/sites/default/files/2022-04/statutoscuolanormalesuperiore2022.pdf. Accessed: 30-09-2025.

Scuola Superiore Sant'Anna. 2018. Statuto della Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna. Accessed: 30-09-2025.

Scuola Superiore Sant'Anna. 2024. Statuto della Scuola Superiore di Studi Universitari e di Perfezionamento Sant'Anna. https://www.santannapisa.it/sites/default/files/2024-01/Statuto%20Scuola%20Superiore%20Sant%E2%80%99Anna.pdf. Accessed: 30-09-2025.

Dagmar Stahlberg, Friederike Braun, Lisa Irmen, and Sabine Sczesny. 2007. Representation of the Sexes in Language. In *Social communication*, Frontiers of social psychology, pages 163–187. Psychology Press, New York, NY, US.

Università Ca'Foscari Venezia. 2013. Statuto di Ateneo. Accessed: 30-09-2025.

Università Ca'Foscari Venezia. 2023. Statuto di Ateneo. https://www.unive.it/pag/fileadmin/user_upload/ateneo/norme_regolamenti/statuto/statuto-2023.pdf. Accessed: 30-09-2025.

Università degli Studi della Tuscia. 2019. Statuto. https://www.unitus.it/wp-content/uploads/2023/10/Statuto-modifica-2019-7.pdf. Accessed: 30-09-2025.

Università degli Studi della Tuscia. 2024. Statuto. https://www.unitus.it/wp-content/uploads/2025/03/STATUTO-2024-impaginato.pdf. Accessed: 30-09-2025.

Università degli Studi di Bergamo. 2012. Statuto. https://www.unibg.it/sites/default/files/ateneo_bergamo_spa/statuto_carta_intestata_20.12.pdf. Accessed: 30-09-2025.

Università degli Studi di Bergamo. 2023. Statuto dell'Università degli Studi di Bergamo. https://www.unibg.it/sites/default/files/media/documents/2023-08-22/1.%20DR%20modifica%20Statuto_0.pdf. Accessed: 30-09-2025.

Università degli Studi di Brescia. 2020. Statuto. https://www.unibs.it/sites/default/files/2021-04/Statuto%20autonomia.pdf. Accessed: 30-09-2025.

Università degli Studi di Brescia. 2024. Statuto. https://www.unibs.it/sites/default/files/2024-06/Statuto_pdf.pdf. Accessed: 30-09-2025.

Università degli Studi di Ferrara. 2018. Prontuario dell'Università degli Studi di Ferrara per l'uso del genere nel linguaggio amministrativo e per la redazione di documenti accessibili. https://www2.unife.it/ateneo/uffici-old/cnapi/coordinamento-politiche-pari-opportunita-e-disabilita/prontuario-dell2019universita-degli-studi-di-ferrara-per-l2019uso-del-genere-nel-linguaggio-amministrativo-e-per-la-redazione-di-documenti-accessibili.

Università degli Studi di Padova. 2017. Generi e linguaggi. Linee guida per un linguaggio amministrativo e istituzionale attento alle differenze di genere. https://www.unipd.it/sites/unipd.it/files/2017/Generi%20e%20linguaggi.pdf. Accessed: 30-09-2025.

Università degli Studi di Palermo. 2019. Statuto dell'Università degli Studi di Palermo. https://www.unipa.it/servizi/prevenzionedellacorruzione/.content/documenti/statuto_e_codici_di_ateneo/Statuto-dell_Ateneo-di-Palermo---14-maggio-2019-in-vigore-dal-12-06-2019.pdf. Accessed: 30-09-2025.

Università degli Studi di Palermo. 2022. Statuto dell'Università degli Studi di Palermo. https://www.unipa.it/servizi/prevenzionedellacorruzione/.content/documenti/statuto_e_codici_di_ateneo/Statuto-dell_Ateneo-di-Palermo---03-giugno-2019---in-vigore-dal-30-06-2022.pdf. Accessed: 30-09-2025.

Università degli Studi di Siena. 2005. Statuto dell'Università degli Studi di Siena. https://www.unisi.it/sites/default/files/allegatiparagrafo/statuto_2005.pdf. Accessed: 30-09-2025.

Università degli Studi di Siena. 2024. Statuto. https://www.unisi.it/sites/default/files/albo_pretorio/allegati/Statuto%20UniSi%202024_definitivo_0.pdf. Accessed: 30-09-2025.

Università degli Studi di Trento. 2012. Statuto. https://r.unitn.it/filesresearch

/images/download/statuto-2012.pdf. Accessed: 30-09-2025.

Università degli Studi di Trento. 2017. Per un uso del linguaggio rispettoso delle differenze. https://www.unitn.it/sites/default/files/2024-11/documentolinguaggio29mar%20%281%29.pdf.

Università degli Studi di Trento. 2024. Statuto. https://www.unitn.it/sites/default/files/2024-11/Statuto2024.pdf. Accessed: 30-09-2025.

Università degli Studi di Trento. 2024a. Regolamento generale di Ateneo. https://www.unitn.it/sites/default/files/2024-10/Regolamento_generale_ateneo2024.pdf. Accessed: 30-09-2025.

Università degli Studi di Trento. 2024b. UniTrento vara il Regolamento di Ateneo. Tutto al femminile. https://pressroom.unitn.it/comunicato-stampa/unitrento-vara-il-regolamento-di-ateneo-tutto-al-femminile. Accessed: 30-09-2025.

Università per Stranieri di Siena. 2020. Linguaggio amministrativo e differenze di genere. Linee guida 2020. Accessed: 30-09-2025.

Università per Stranieri di Siena. 2021. Statuto. https://www.unistrasi.it/public/articoli/317/Files/Estratto%20Pdf%20gazzetta%20ufficiale%20n.%20117%20-%2018..05.2021.pdf. Accessed: 30-09-2025.

Università per Stranieri di Siena. 2024. Statuto. https://www.unistrasi.it/public/articoli/317/Files/GU%20n.%20136%2012%20giugno%202024%20Statuto.pdf. Accessed: 30-09-2025.

Emma Zanoli. 2022. A Linguistic Perspective on Feminisation and Gender-Inclusive Language: How Partitive Constructions Can Be Used to Uncover Implicit Gender Bias in Italian. *Annali di Ca' Foscari. Serie occidentale*, (56).

Giuseppe Zarra. 2023. Ideologie linguistiche e nomi femminili di professioni e di cariche. *Circula*, (17-18):276–298.