Valicə: A gender-neutralized dataset of learner Italian

Céline Leuzinger* **Caroline Grand-Clement**
Ricardo Muñoz Sánchez **Tom Södahl Bladsjö**
celine.leuzinger@gu.se
Gothenburg University, Sweden

Abstract

Investigations of gender-neutral and non-binary language in Romance languages often fail to take into account learner language. There is a lack of material to study how grammatical error correction models handle non-binary language which uses new forms or morphemes. In this paper, we offer Valicə, a gender-neutralized version of the Valico treebank, containing Italian learner texts and their corrections. We present the procedure to gender-neutralize Italian sentences, both grammatically correct and incorrect. We implement two different systems to identify which tokens should be neutralized: a rule-based system and a machine learning system, both of which prove equally efficient. The token swap is rule-based, using the neomorphemes -ə and -ɜ to signal singular and plural gender-neutrality. The Valicə dataset can be used to evaluate or train language learning applications for learners of Italian.

Gendered version
C'era anche nel parco un giovane ragazzo.
There was also in the parc a young boy.

Gendered version (corrected)
C'era anche un giovane ragazzo nel parco.
There was also a young boy in the parc.

Gender-neutralized version
C'era anche nel parco unə giovane ragazzə.
There was also in the parc a young person.

Gender-neutralized version (corrected)
C'era anche unə giovane ragazzə nel parco.
There was also a young person in the parc.

Figure 1: An example sentence from the Valicə dataset.

1 Introduction

Italian, as most other Romance languages, has two grammatical genders: feminine and masculine. While inanimate objects have an opaque grammatical gender, that of animate nouns must concur, if possible, with the social gender of the person it refers to. The gender of a person being spoken about must always be specified, through the nouns used to refer to them but also through the adjectives, participles, pronouns and determiners linked to them, with masculine being the default. This can lead to biases, from both a cognitive or computational point of view (Gygax et al., 2021; Devinney, 2024).

The gendering in Romance languages also creates difficulties for people with non-binary gender identities. The term *non-binary* can be understood as an umbrella term encompassing gender identities that do not fit into frameworks that view gender as a binary between man and woman (Cordoba, 2023b). The fact that grammatical gender in Romance languages traditionally allows no options outside of the masculine/feminine binary makes it hard to correctly refer to individuals who do not fit into that binary. Meanwhile, being misgendered (i.e. being referred to by gendered terms that do not correspond to your gender identity) can be very distressing. Repeated experience of misgendering has been shown to have a negative effect on mental health (McLemore, 2015), leading to symptoms of anxiety and depression (Puckett et al., 2023). Various types of gender neutralization, such as rephrasing sentences to not have to specify gender, can be a way to both counteract the masculine default bias, and provide ways to talk about non-binary people without misgendering. However, queer communities have also come up with creative ways to use language to be able to talk about their own experiences, such as specifically non-binary word forms (see Section 2.1).

*Corresponding author

Proceedings of the 7th Workshop on Gender Bias in Natural Language Processing (GeBNLP), pages 7–20
November 9, 2026 ©2026 Association for Computational Linguistics

In NLP research, gender is almost exclusively framed as a binary (Devinney et al., 2022). Despite that, there are multiple resources and datasets that have been released for French, Italian, and Swedish (see Section 2.3). However, most research on non-binary language focuses on text that is assumed to be written by first language speakers (see Section 2), thereby overlooking other categories of speakers, such as second language learners. Language learning technologies favor a standardized version of the target language, which risks silencing non-standard usage of the language meant to express one's identity. We posit that language learners should be able to express their gender identity in their target language.

In this paper we introduce a system that can recognize gendered tokens in Italian and neutralize them to a non-binary equivalent. It consists of two stages: the first is the identification of human animate terms and their coreferents, executed either with a rule-based model or a machine learning model; the second is the gender-neutralization of the identified terms with a rule-based algorithm. The whole dataset creation pipeline is illustrated in Figure 2. These systems can be used to find and neutralize gendered tokens in Italian datasets annotated with Universal Dependencies (UD).[1]

We present Valicə, a dataset of Italian learner data along with correction hypotheses,[2] that includes non-binary neoforms.[3] This dataset is meant to evaluate whether language learning systems, such as grammatical error correction tools, accept non-binary language. The dataset can also be used to fine-tune said systems to ensure language learners can express their gender identities in their target language without being censored by the tools that are meant to help them.

We make all our data and code available in a GitHub repository.[4]

[1]UD is a cross-linguistic framework of morphosyntactic annotation, which marks part of speech, morphological features (such as gender, number) and syntactic features (such as head and dependent relationships). The current version of the framework was introduced by de Marneffe et al. (2021).

[2]Each learner sentence is accompanied by a corrected version, which remains as faithful to the original as possible. We call these "correction hypotheses" to reflect the fact that there is not necessarily a unique "correct" version of each text.

[3]Non-binary neoforms are forms of gendered words which adopt a new and non-binary declension. See Section 4.2 for a description of the used forms.

[4]https://github.com/rimusa/Valic-A-gender-neutralized-dataset-of-learner-Italian

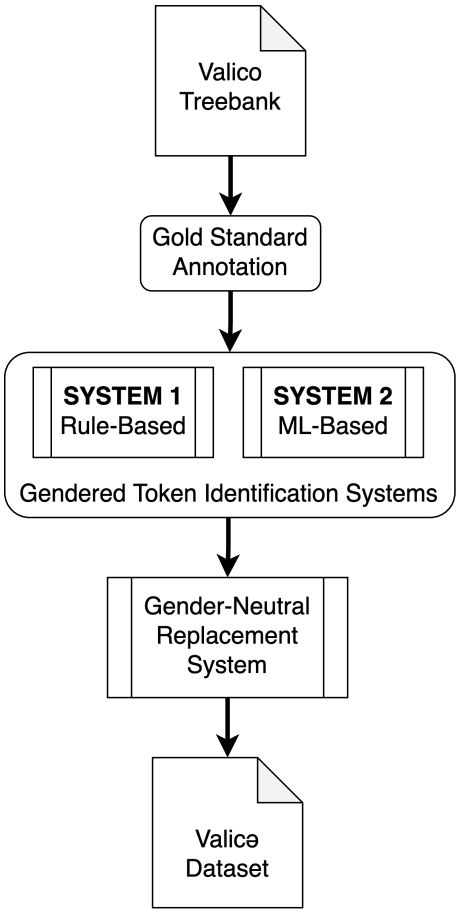


Figure 2: Pipeline for generating the non-binary version of the Valico treebank.

2 Background

2.1 Gender neutrality and non-binary language

Traditional conceptions of gender in Western countries often treat it as a man/woman binary (Cordoba, 2023a). However, gender can be conceptualized more broadly, as a reflection and expression of personal identity created through repetition within a social and cultural context (West and Zimmerman, 1987). Fontenele and Abi Zeid Daou (2025) note that gender identities beyond the binary have existed throughout the ages in different cultures across the world.

Grand-Clement (2026) points out that language plays an important role in the construction of non-binary identities. To avoid a binary masculine/feminine gendering, non-binary people can turn to gender-neutral or non-binary language. Gender-neutral language aims to conceal or eliminated gendered information and can also be used to reduce gender bias at large, while non-binary language

escapes the masculine/feminine gender binary explicitly and can be used to refer to non-binary people specifically. Some strategies for gender-neutrality involve removing gender specifications where possible, sometimes significantly changing the sentence structure (Savoldi et al., 2025). However, the strategies used to create gender-neutral language can also overlap with those used to create non-binary language. These overlapping strategies often rely on non-standard usage of language, also known as *queering* the language (Milani, 2018). In this paper, we view non-binary language as a subtype of gender neutral language, and our modification of the Valico dataset as a type of gender-neutralizaiton.

For instance, pronouns can be used to reflect gender-neutral and/or non-binary language by expanding on the use cases of previously existing pronouns (e.g. singular *they* in English) or using new ones altogether (also known as neopronouns). In the case of languages with feminine/masculine grammatical gender agreement, such as German or Romance languages, new declensions are often made to allow for gender-neutral or non-binary forms of adjectives, nouns, participles, and any other words that require gender agreement. We present how this works in the case of Italian in the following section.

2.2 Non-binary language in Italian

Grammatical gender in Italian presents two forms, masculine and feminine, which most often reflect a person's social gender when referring to people. The language requires gender agreement between nouns, pronouns, adjectives, articles and certain cases of participles (Nodari, 2024). Typical endings for masculine are -o in singular and -i in plural, while feminine endings are -a in singular and -e in plural. Some forms are epicene, meaning they can be used for masculine or feminine. This is the case, for example, of adjectives ending in -e in the singular (in plural, these always end in -i), such as *grande* (big).

To construct non-binary forms in Italian, common practice is to replace the final vowel. What takes its place varies, from alternative vowels (-u), to typographical symbols like the star (*), to the IPA symbols ə and ɜ, now used more commonly, for singular and plural respectively (Baiocco et al., 2023). Different strategies can be used for epicene forms, which can remain the same or adopt an alternative ending. For words that have a different stem in the masculine and feminine (such as *attore* (actor) and *attrice* (actress)), alternative versions may keep the masculine or feminine stem when neutralized, yielding *attorə* or *attricə*, though keeping the masculine stem is often proposed by online resources such as Italiano Inclusivo[5] or Gender in Language[6].

2.3 Gender-neutralization in NLP

Several datasets have been created to explore how NLP systems respond to gender-neutral or non-binary language. We briefly present these datasets throughout this section, as well as some approaches for gender-neutralization.

Flesch and Burnett (2025) put together a French corpus of first-person statements from social media annotated for gender, including a non-binary category, for gender inference.

In Swedish, much like English, a person's gender is revealed mostly through third-person pronouns. Björklund and Devinney (2023) investigated POS tagging when confronted with the gender-neutral pronoun *hen*. For this, they created and made available an algorithm to 'en*hen*ce' Swedish corpora - that is, to swap gendered pronouns for *hen*.

Grand-Clement (2026) presented ENBYS, a parallel dataset of gender-neutralized sentences across English, Polish and French, adapted from the Europarl corpus (Koehn, 2005), to evaluate machine translation on the recognition and generation of gender-neutral or non-binary language.

GATE (Rarrick et al., 2023) is a machine translation benchmark for evaluating gender bias. It includes gender-ambiguous sentences in English and their possible translations in Italian, translated either into feminine or masculine. Piergentili et al. (2024) created NEO-GATE, an augmented dataset that also includes gender-inclusive alternatives for each translation, based on neomorphemes. Their dataset used tags that mark gendered morphemes or full tokens (such as determiners, possessives) which can be swapped out to the desired neomorpheme, allowing a number of strategies for gender-neutralizing languages (see Section 2.2).

Work has also been done on using LLMs for gender-neutral rewriting and the creation of dataset, for example for Polish (Wróblewska and Żuk, 2025) or Italian (Piergentili et al., 2025). However, their gender-neutralization sometime still operated within a binary framework. Piergentili et al.

[5] https://italianoinclusivo.it/

[6] https://www.genderinlanguage.com/italian

(2025)'s gender-neutralization for Italian is based on the elimination of gendered expressions, and they explicitly excluded non-standard language techniques such as the use of the schwa when prompting the LLMs. In contrast, our models aim to adopt a gender-neutralized Italian by using explicitly non-binary (and so, often non-standard) forms.

3 Gendered Token Identification

The first step in creating a non-binary dataset from binary-gendered sentences is to identify which tokens mark gender and thus need to be neutralized. To this end, we devise and compare two systems: one which uses rule-based identification and another that is based on machine learning, using Zmigrod et al. (2019)'s procedure.

3.1 The Valico treebank

We base our dataset on the Valico treebank (Di Nuovo et al., 2022), which consists of 398 sentences written by L2 Italian learners and their corresponding correction hypotheses given by teachers. The Valico treebank is annotated for error types as well as UD.

3.2 Annotation of gold label

We annotate the Valico Treebank for human animate gendered tokens, as well as their dependents and coreferents that agree in gender. The annotated dataset represents the gold standard against which we compare our two gendered token identification systems. The annotation process is done fully manually by one of the authors, using SACR, a coreference chain annotation tool (Oberle, 2018).

Human animate nouns include:

- Human animate nouns, such as *lettore* (reader), *attrice* (actress), *bambini* (children).
- Proper nouns, such as *Marco*, *Giulia*, *Clark*.
- Pronouns like *lui* (he), *lei* (she), *Lei* (polite 3rd person form) that refer to a human animate.

Adjectives, past participles, determiners and pronouns that require gender agreement with these (pro)nouns are also marked. Adjectives and pronouns that do not vary in gender, such as *forte* (strong) or *gliene* (contraction of the 3rd person dative pronoun *gli* and the pronoun *ne* (about it)) are not included in the gendered tokens.

Gender agreement error
*[...] lui era il suo **ragazza** [...]*
[...] he was her **girlfriend** [...]

Gender form error
*[...] **uno** uomo che si chiama Matteo stava leggendo*
[...] a man named Matteo was reading.

Figure 3: An example of gender errors from the Valico dataset

3.3 Filtering of gender errors

We create two versions of Valicə: a filtered version, where sentences containing gender errors are not neutralized, and an unfiltered version, where those sentences are neutralized. The filtered version aims to preserve the quality of the learner texts and error annotations: the process of gender neutralization can indeed modify the nature of gender errors, creating potential mismatches with existing error annotations. This version of the dataset can be used to check if downstream applications, such as grammatical error checking systems, accept non-binary forms when not part of an error span. On the other hand, the unfiltered version gender-neutralizes tokens even if they are part of a gender error span. This version of the dataset can be used to check whether grammatical error checking systems are capable of detecting errors involving non-binary forms. An example of each gender error type (gender agreement, gender form) is shown in Figure 3.

The unfiltered dataset contains 398 sentences, and the filtered dataset contains 299 sentences.

3.4 Human animate noun list

We craft a list of human animate nouns that serves as an anchor to find gendered tokens in the base dataset. To craft the list, we extract all the unique nouns and proper nouns from the Valico treebank, and exclude all nouns that are not human animates (such as *cane* (dog), *muscoli* (muscles), *Barcelona*).

3.5 System 1: Rule-based identification of gendered tokens

The list of human animate nouns serves as the anchor to find gendered tokens according to the following process:

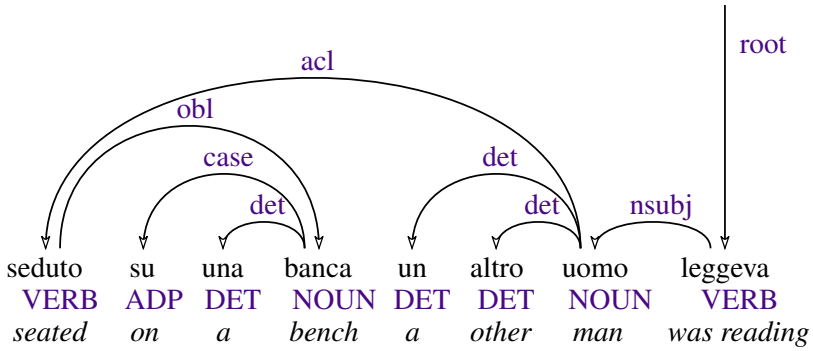


Figure 4: UD annotation of the phrase *seduto su una banca un altro uomo leggeva*, with translation.

- **Step 1**: the system identifies human animate nouns based on the hand crafted list and extracts their relevant morphosyntactic features: form, number, dependents, head and the type of dependency relation with the head.

- **Step 2**: the system checks whether the dependent tokens are relevant for gender neutralization. A dependent token is considered relevant if it agrees with the gender of its head. Whether a token agrees with the gender of its head depends on both its part-of-speech and the nature of the dependency relation with the head. Only adjectives, nouns, verbs, determiners and pronouns can potentially be relevant dependent tokens of a human animate noun. For example, in the phrase *seduto su una banca un altro uomo leggeva* (seated on a bench another man was reading), illustrated in Figure 4, the human animate anchor is *uomo* (man), and has *un* (a), *altro* (other) and *seduto* (seated) as dependents, all of which agree with *uomo* (man). These are therefore extracted as relevant for gender neutralization. The conditions for relevance of dependent tokens are explicited in Table 1.

- **Step 3**: the system checks whether the head of the human animate anchor is relevant for gender neutralization. This depends on both the part-of-speech of the head and the nature of the dependency relationship with the anchor. In the example above, *leggeva* (was reading) is the head of *uomo* (man). However, the verb does not agree with the gender of its head, as there is no gender marking on verb. It is therefore not extracted. The conditions for relevance of head tokens are explicited in Table 1.

- **Step 4**: the system extracts the relevant tokens as well as their morphosyntactic features.

3.6 System 2: Machine learning-based identification of gendered tokens

Our machine learning-based model is an adaptation of Zmigrod et al. (2019)'s process for counterfactual data augmentation. Using a Markov random field (MRF), the original process converts animate nouns within a sentence from masculine to feminine, or vice-versa, and reinflects the sentence as appropriate. It operates in three steps: first, an algorithm identifies the animate nouns and replaces them with the alternative gender; second, an inference model infers the morphological tags for each token that should be reinflected; third, a reinflection model replaces the tokens for which the tags are different from the original sentence's to produce the appropriate forms.

We follow their approach as it is easily applicable to our UD-annotated dataset. As our aim is to create sentences with non-binary forms and there is no such data to train a reinflection model, we keep only the first two steps, with some modifications. Our "reinflection" step is instead carried out by our rule-based replacement algorithm, presented in Section 4.

To identify and replace animate nouns, we use our list of human animate and proper nouns, and modify the code to check both nouns and proper nouns.

For the second step, we use Zmigrod et al. (2019)'s code to train a neural MRF for Italian. The data used to train the model is a mix of the Valico treebank itself, both learner and corrected data, and the Italian part of ParTUT (Sanguinetti and Bosco, 2015), an Italian-English-French parallel treebank covering different genres and domains. This led to a total of 2488 sentences, with an 80/10/10% split

POS	Relevance conditions for dependent tokens	Relevance conditions for head tokens
any	—	The dependency relation is either {nsubj, nsubj:pass, csubj}
ADJ	The token takes gender marking	The token takes gender marking
NOUN	The dependency relation is "appositional modifier"	Any noun
VERB	The verb form is "participle", the preceding auxiliary is not *avere* (to have) and the dependency relation is either {acl, acl:relcl, xcomp, cop}	The verb form is "participle" and preceding auxiliary is not *avere* (to have)
PRON & DET	The pronoun type is either {relative, interrogative} and the lemma starts with "qu", the pronoun type is not {negative, exclamative}	—

Table 1: Conditions of relevance for dependent and head tokens. For example, an adjective that is a dependent token of a human animate anchor is only relevant if it takes gender marking. Similarly, a verb that is the head token of a human animate anchor is only relevant if it is conjugated in the past participle, and not preceded by the auxiliary *avere* (to have). The UD dependecy relations are described on the UD website.[7]

for training, development and test set.

The original code assumes a binary masculine/feminine replacement. We use this code along with our trained MRF model to infer the new morphological tags after the animate nouns are swapped. This way we can identify which other tokens need to be updated. It is important to note that we do not care about the new gender morphological tags these tokens are given, but rather that they are different than their original tag.

Finally, rather than reinflecting the tokens with the new given tags, we pass all tokens that need to be changed through our rule-based replacement algorithm described in Section 4.

3.7 Results and discussion

The results of our gendered token identification models on the annotated version of the Valico treebank are reported in Table 2 for the learner sentences and Table 3 for the correction hypotheses, for both unfiltered and filtered settings. The results are based on gendered tokens (positive predictions) only.

3.7.1 Performance across systems

Both systems show high accuracy and precision across settings (i.e. learner texts / correction hypotheses, filtered / unfiltered sentences). On the other hand, recall is lower across systems and settings (0.69-0.73). This similar performance could be due to the fact that both systems have the same anchor: the handcrafted list of human animate nouns. However, minor differences can also be noted: the rule-based model displays slightly higher precision, while the ML model achieves higher recall.

The fact that the rule-based system shows higher precision is unsurprising: manually crafting rules ensures their relevance and quality. The system is therefore fully explainable and can be scaled to cover other gender agreement cases, in Italian or another language. However, such scaling requires expertise in both UD annotation and gender agreement. Another limitation is that it cannot find relevant gendered tokens if not morphosyntactically related to a human animate. For example, in the sentence "I saw a *woman* and *she* was buying a book", the model will not identify the pronoun "she" as relevant.

On the other hand, the machine learning-based model can recognize coreferents even if synctactically not related, albeit not in a consistent manner. This might explain the higher recall on the filtered dataset, as it can potentially detect coreferences that the rule-based model misses. However, we also note certain limitations, such as inconsistencies in the flagging of certain gendered tokens due to the probabilistic nature of the model: for exam-

[7] https://universaldependencies.org/u/dep/index.html

System	Acc.	Precision	Recall	F1-score
Rule-based (filtered)	**0.95**	**0.99**	0.72	0.83
ML-based (filtered)	0.94	0.96	**0.73**	0.83
Rule-based (unfiltered)	0.94	**0.99**	0.69	0.81
ML-based (unfiltered)	0.94	0.95	**0.72**	**0.82**

Table 2: Accuracy, Precision, Recall and F1-score for the gendered token identification of the learner texts. Precision, Recall and F1-scores are based on positive predictions (i.e. gendered tokens) only.

System	Acc.	Precision	Recall	F1-score
Rule-based (filtered)	0.95	**0.99**	0.72	0.83
ML-based (filtered)	0.95	0.96	**0.73**	0.83
Rule-based (unfiltered)	0.94	**0.98**	0.68	0.81
ML-based (unfiltered)	0.94	0.97	**0.73**	**0.83**

Table 3: Accuracy, Precision, Recall and F1-score for the gendered token identification of the corrected texts. Precision, Recall and F1-scores are based on positive predictions (i.e. gendered tokens) only.

ple, the identification of participles as gendered tokens is inconsistent, as the model occasionally misses gendered participles.

4 Gender-Neutralizing the Data

To create non-binary data, the tokens detected by the models must be swapped for their non-binary equivalent.

Non-binary forms in Italian are a recent trend and vary considerably depending on the individuals, and there are few set rules for non-binary language. We therefore base our decisions on either:

(a). Popular online resources, such as Italiano Inclusivo[8] and the community-based platform Pronom.it[9]

(b). Information collected from a voluntary non-binary native speaker of Italian.

(c). The semantic and morphological features of the token, particularly in cases where several non-binary alternatives are possible.

Our final set of replacement rules contains a mix of individual replacements and rule-based replacements. Individual replacements are one-to-one correspondences between a token and its non-binary equivalent, while rule-based replacements apply to an entire category of token.

[8] https://italianoinclusivo.it/ (Accessed: 2026-07-27)

[9] https://pronom.it/ (Accessed: 2026-07-27)

4.1 Individual replacements

Individual replacements concern the majority of POS: human animate nouns, proper nouns, determiners and pronouns.

We craft a list of one-to-one correspondence between each form and its non-binary equivalent, as in the examples in Table 4.

A number of nouns and proper nouns have several non-binary alternatives depending on their morphological and semantic gender behaviour. These cases are spelled out below, and we detail the choice made for each one:

- For human animate nouns that change their stem between the feminine and masculine form, we take the masculine form and replace only the morphological suffix with -ə (singular) or -ɜ (plural), following the recommendation of our native speaker informant. Example: *attore* (actor) and *attrice* (actress) become *attorə*.
- For human animate nouns that do not take any morphological gender marking, we leave the base form. Example: *delinquente* (deliquent) stays *delinquente*.
- For human animate nouns that have morphological gender and semantic gender, such as *fratello* (brother) and *sorella* (sister), but no commonly accepted non-binary form, we keep the base form and replace only the morphological suffix with -ə (singular) or -ɜ

POS	Form	
	Gendered	Non-binary
NOUN	adulto	adultə
PROPN	Giulio	Giulio
DET	tua	tuə
PRON	lui	ləi

Table 4: Example of handcrafted one-to-one correspondence. The full list is provided in Appendix A.

(plural). Example: *sorella* (sister) becomes *sorellə*, *fratello* (brother) becomes *fratellə*. On the other hand, *uomo* (man) and *donna* (woman) have a commonly used alternative, *persona* (person). We therefore use its gender-neutralized form *personə* as the non-binary equivalent of *uomo/donna*.

- For proper nouns, we leave the base form. Example: *Marco* stays *Marco*.
- For pronouns and determiners positioned in front of a word starting with a vowel, a possible non-binary alternative to the apostrophe (*un'*, *qual'*, *l'*) is the star (*un**, *qual**, *l**). However, this creates ungrammatical results in phrases such as *l'uomo* (the man). In this case, the phrase is gender-neutralized to *l*personə* (the person), which is ungrammatical. We therefore made the choice to also add the morpheme -ə or -ɜ at the end of the token to create a non-binary variant of these pronouns and determiners. Example: *l'uomo* (the man) becomes *lə personə* (the person).

4.2 Rule-based replacements

For the remaining parts of speech, adjectives and verbs, we use a rule-based system for replacement:

- If the token's "number" feature is "singular" and the last letter is a vowel, replace the last letter with -ə.
- If the token's "number" feature is "plural" and the last letter is a vowel, replace the last letter with -ɜ.
- In all other cases, keep the base form.

The rules above therefore handle adjectives that are marked for gender, such as *bella* / *bello* (beautiful) as well as past participles that take gender agreement with the verb, such as *arrivata* / *arrivato* (arrived).

We gender-neutralize the output of both the rule-based model and the machine learning-based model. Both variants are available on our Github repository.

4.3 Final sentence extraction

We adapt Zmigrod et al. (2019)'s script to extract the sentences from the output CoNLL-U files in order to adequately extract the gender-neutralized sentences, in particular to reconstitute articulated prepositions in the correct form.

To do so, we go through every token and add their forms to our output sentence. If a token is a multitoken, we check whether one of its parts has been swapped. If that is the case, we look up the equivalent non-binary reconstituted token in the hand crafted list of articulated prepositions shown in Appendix A and add it to the sentence.

After the sentences are extracted, we conduct manual quality checks to control for any missed token or errors that might have occurred during the neutralization process, and correct them in our final dataset. The result is Valicə, a gender-neutralized dataset of learner Italian. The statistics for Valicə, including the number of neutralized sentences and tokens, can be found in Table 5 for the non-corrected version and Table 6 for the corrected version.

4.4 Dataset statistics

Learner texts: in the filtered setting, 197 sentences are neutralized for the rule-based model, yielding a total of 632 non-binary tokens. For the machine learning-based model, 195 sentences are neutralized, for a total of 662 non-binary tokens.

In the unfiltered setting, 271 sentences are neutralized for the rule-based model against 297 for the machine learning-based model. This yields a total of 918 tokens for the rule-based system and 1000 for the machine learning-based system.

Corrected texts: in the filtered settings, 194 sentences are gender neutralized for the rule-based model, yielding a total of 632 non-binary tokens. The machine learning-based model neutralizes a total of 189 sentences and 662 non-binary tokens.

In the unfiltered setting, the number of sentences rises to 266 for both models, yielding 913 non-binary tokens for the rule-based system and 989 for the machine learning-based model.

System	N. neutralized tokens	N. neutralized sentences
Rule-based (filtered)	632	197
ML-based (filtered)	662	195
Rule-based (unfiltered)	918	271
ML-based (unfiltered)	1000	297

Table 5: Dataset statistics for Valicə.

System	N. neutralized tokens	N. neutralized sentences
Rule-based (filtered)	632	194
ML-based (filtered)	662	189
Rule-based (unfiltered)	913	266
ML-based (unfiltered)	989	266

Table 6: Dataset statistics for the corrected Valicə.

5 Conclusion

In this paper we compare two different systems to identify gendered tokens in Italian text. While their performance is similar, the rule-based system has higher precision and the ML-based system slightly higher recall. We also present a gender neutralization system and use it to create a non-binary version of the Valico treebank of Italian learner texts, which we dub Valicə.

Although we focus our work on language learner texts, the systems introduced in this paper can also be used for texts written by native speakers. Future work could compare how our systems perform across different domains and registers, with both L1 and L2 texts.

Another way of expanding this work would be via community involvement throughout the design and implementation process. This would allow us to further refine our work and create several versions of Valicə, thereby reflecting the variety of non-binary identities. Language is an essential tool for expressing one's identity, especially for trans and non-binary communities, and it is important to ensure that they are part of the conversation and properly represented on and in their own terms.

Limitations

The procedure to gender-netutralize Italian sentences offered in this paper has limitations, both at the level of the token identification and at that of the token swapping itself.

Both identification systems presented in Section 3 rely on UD annotation, which puts constraints on the input data one can use. The starting point for token identification, in both systems, is human animacy and proper nouns, meaning we only identify third-person referents to neutralize. The list of human animates and proper nouns used is crafted manually, a task which is not easily scalable.

Each of the systems also has its own limitations. System 1 requires expert linguistic knowledge due to the handcrafted rules, which can make future-proofing or adaptation more time-consuming and difficult. For System 2, the addition of the proper nouns as anchor points means that gender feature tags have to be added for names to the input files, making it difficult to scale up.

In regards to the gender-neutralization, we make decisions about the rules to implement, which represent only one way of making gender-neutral or non-binary language in Italian.

Ethical Considerations

This paper deals with how people use language to express their identities in ways that diverge from the norm. However, the ways in which people use gender-neutral and/or non-binary language can vary greatly both through time and across different groups and individuals (Lardelli and Gromann, 2023). As such, we are aware that we risk presenting this kind of language as if it were monolithical, rather than an ever-evolving way of speaking.

References

Roberto Baiocco, Fau Rosati, and Jessica Pistella. 2023. Italian proposal for non-binary and inclusive language: The schwa as a non-genderspecific ending. *Journal of Gay & Lesbian Mental Health*, 27(3):248–253.

Henrik Björklund and Hannah Devinney. 2023. Computer, enhence: POS-tagging improvements for non-binary pronoun use in Swedish. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 54–61, Varna, Bulgaria. INCOMA Ltd., Shoumen, Bulgaria.

Sebastian Cordoba. 2023a. Becoming non-binary: Language and identity. In *Non-Binary Gender Identities*, first edition, pages 1–27. Routledge.

Sebastian Cordoba. 2023b. *Non-binary gender identities: the language of becoming*. Gender and sexualities in psychology. Routledge, Taylor & Francis Group.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2):255–308.

H Devinney. 2024. *Gender and representation: investigations of bias in natural language processing*. Report / UMINF. Department of Computing Science, Umeå University.

H Devinney, Jenny Björklund, and Henrik Björklund. 2022. Theories of gender in NLP bias research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '22, pages 2083–2102, New York, United States. Association for Computing Machinery.

Elisa Di Nuovo, Manuela Sanguinetti, Alessandro Mazzei, Elisa Corino, and Cristina Bosco. 2022. Valico-ud: Treebanking an italian learner corpus in universal dependencies. *IJCoL. Italian Journal of Computational Linguistics*, 8(8-1).

Marie Flesch and Heather Burnett. 2025. Characterizing non-binary French: A first step towards debiasing gender inference. In *Proceedings of the 6th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, pages 160–170, Vienna, Austria. Association for Computational Linguistics.

Rodrigo Fontenele and Margarita Abi Zeid Daou. 2025. Transgender history, part I: An anthropology of gender-nonconformity across ages and cultures. *Behavioral Sciences & the Law*, 43(3):323–328.

Caroline Grand-Clement. 2026. *Presenting ENBYS The Europarl Non-BinarY Sentences: an English-Polish-French Dataset for Machine Translation Evaluation of Non-Binary Gender Inclusion*. University of Gothenburg.

Pascal Gygax, Sayaka Sato, Anton Öttl, and Ute Gabriel. 2021. The masculine form in grammatically gendered languages and its multiple interpretations: a challenge for our cognitive system. *Language Sciences*, 83:101328.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of Machine Translation Summit X: Papers*, pages 79–86, Phuket, Thailand.

Manuel Lardelli and Dagmar Gromann. 2023. Translating non-binary coming-out reports: Gender-fair language strategies and use in news articles. *The Journal of Specialised Translation*, 2023(40):213–240.

Kevin A. McLemore. 2015. Experiences with misgendering: Identity misclassification of transgender spectrum individuals. *Self and Identity*, 14(1):51–74.

Tommaso M. Milani. 2018. *Queering language, gender and sexuality*. Equinox Publishing Ltd.

Rosalba Nodari. 2024. Gender-inclusive strategies in italian: stereotypes and attitudes. In Falco Pfalzgraf, editor, *Public Attitudes Towards Gender-Inclusive Language: A Multilingual Perspective*, pages 243–286. De Gruyter Mouton.

Bruno Oberle. 2018. SACR: A drag-and-drop based tool for coreference annotation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Andrea Piergentili, Beatrice Savoldi, Matteo Negri, and Luisa Bentivogli. 2024. Enhancing gender-inclusive machine translation with neomorphemes and large language models. In *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (Volume 1)*, pages 300–314, Sheffield, UK. European Association for Machine Translation (EAMT).

Andrea Piergentili, Beatrice Savoldi, Matteo Negri, and Luisa Bentivogli. 2025. Gender-neutral rewriting in Italian: Models, approaches, and trade-offs. In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, pages 899–910, Cagliari, Italy. CEUR Workshop Proceedings.

Jae A. Puckett, Christina Dyar, Meredith R. Maroney, Brian Mustanski, and Michael E. Newcomb. 2023. Daily experiences of minority stress and mental health in transgender and gender-diverse individuals. *Journal of Psychopathology and Clinical Science*, 132(3):340–350.

Spencer Rarrick, Ranjita Naik, Varun Mathur, Sundar Poudel, and Vishal Chowdhary. 2023. GATE: A challenge set for gender-ambiguous translation examples. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, pages 845–854. Association for Computing Machinery.

Manuela Sanguinetti and Cristina Bosco. 2015. Part-TUT: The turin university parallel treebank. In Roberto Basili, Cristina Bosco, Rodolfo Delmonte, Alessandro Moschitti, and Maria Simi, editors, *Harmonization and Development of Resources and Tools for Italian Natural Language Processing within the PARLI Project*, volume 589 of *Studies in Computational Intelligence*, pages 51–69. Springer International Publishing.

Beatrice Savoldi, Giuseppe Attanasio, Eleonora Cupin, Eleni Gkovedarou, Janiça Hackenbuchner, Anne Lauscher, Matteo Negri, Andrea Piergentili, Manjinder Thind, and Luisa Bentivogli. 2025. Mind the inclusivity gap: Multilingual gender-neutral translation evaluation with mGeNTE. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 13698–13720, Suzhou, China. Association for Computational Linguistics.

Candace West and Don H. Zimmerman. 1987. Doing gender. *Gender and Society*, 1(2):125–151.

Alina Wróblewska and Bartosz Żuk. 2025. Integrating gender inclusivity into large language models via instruction tuning. *Preprint*, arxiv:2508.18466 [cs.CL].

Ran Zmigrod, Sabrina J. Mielke, Hanna Wallach, and Ryan Cotterell. 2019. Counterfactual data augmentation for mitigating gender stereotypes in languages with rich morphology. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1651–1661, Florence, Italy. Association for Computational Linguistics.

A Non-binary Equivalents

Gendered	Non-binary	Gendered	Non-binary
adulti	adultɜ	giovanotto	giovanottə
adulto	adultə	impiegata	impiegatə
amante	amante	innamorati	innamoratɜ
amiche	amicɜ	lettori	lettorɜ
amici	amicɜ	marito	maritə
amica	amicə	moglie	mogliə
amico	amicə	pazo*	pazə
attore	attorə	pazzo	pazzə
attori	attorɜ	persona	personə
attrice	attorə	personaggi	personaggiɜ
bambina	bambinə	personna*	personnə
bambino	bambinə	prince	princə
bambini	bambinɜ	principe	principə
barbaro	barbarə	principessa	principə
carabinieri	carabinierɜ	protagonista	protagonistə
carabiniere	carabiniere	ragazza	ragazzə
carrabinieri*	carrabinieriɜ	ragazze	ragazzɜ
carrabiniere*	carrabiniere	ragazzi	ragazzɜ
collega	collegə	ragazzo	ragazzə
delinquente	delinquente	raggazzo*	raggazzə
dio	diə	rapinato	rapinatə
disparata	disparatə	rapitore	rapitorə
disperata	disperatə	samaritano	samaritanə
donna	personə	saramitano*	saramitanə
dona*	personnə	signorina	signorinə
donne	personɜ	sorella	sorellə
dottore	dottorə	fratello	fratellə
eroe	eroe	stronzo	stronzə
fidanzati	fidanzatɜ	studente	studentə
fidanzato	fidanzatə	tizio	tiziə
fidanzata	fidanzatə	uomo	personə
figlio	figliə	uoma*	personnə
filgia	figliə	uome*	personnə
gente	gente	uomini	personɜ
gentleman	gentleman		

Table 7: Gendered and non-binary forms for common nouns. Forms marked with an asterisk (*) represent misspellings from the original dataset.

Singular			Plural		
Feminine	**Masculine**	**Neutral**	**Feminine**	**Masculine**	**Neutral**
mia	mio	miə	mie	miei	miɜ
tua	tuo	tuə	tue	tuoi	tuɜ
sua	suo	suə	sue	suoi	suɜ
nostra	nostro	nostrə	nostre	nostri	nostrɜ
vostra	vostro	vostrə	vostre	vostri	vostrɜ
loro	loro	loro	loro	loro	loro

Table 8: Gendered and non-binary forms for possessives.

Singular			Plural		
Feminine	**Masculine**	**Neutral**	**Feminine**	**Masculine**	**Neutral**
la	il	lə	le	i	lɜ
la	lo	lə	le	gli	lɜ
l’	l’	lə	le	gli	lɜ

Table 9: Gendered and non-binary forms for definite articles.

Singular		
Feminine	**Masculine**	**Neutral**
una	un	unə
un’	uno	unə

Table 10: Gendered and non-binary forms for indefinite articles.

Plural		
Feminine	**Masculine**	**Neutral**
delle	dei	dɜi/dellɜ
delle	degli	dɜi/dellɜ
alcune	alcuni	alcune

Table 11: Gendered and non-binary forms for partitive articles.

Singular			Plural		
Feminine	**Masculine**	**Neutral**	**Feminine**	**Masculine**	**Neutral**
la	lo	lə	le	li	lɜ
lei	lui	ləi	loro	loro	loro
l	l’	lə	–	–	–
La	–	Lə	–	–	–
Lei	–	Lei	–	–	–

Table 12: Gendered and non-binary forms for object subject pronouns.

Singular			Plural		
Feminine	**Masculine**	**Neutral**	**Feminine**	**Masculine**	**Neutral**
questa	questo	questə	queste	questi	questɜ
quella	quello	quellə	quelle	quelli	quellɜ
quella	quel	quellə	–	–	–
quest’	quest’	questə	–	–	–
quell’	quell’	quellə	–	–	–

Table 13: Gendered and non-binary forms for demonstrative pronouns.

Singular			Plural		
Feminine	**Masculine**	**Neutral**	**Feminine**	**Masculine**	**Neutral**
quale	qual	qualə	quali	quali	qualɜ

Table 14: Gendered and non-binary forms for relative interrogative pronouns.

Singular					Plural			
Feminine	**Masculine**			**Neutral**	**Feminine**	**Masculine**		**Neutral**
la	il	lo	l’	lə	le	i	gli	lɜ
della	del	dello	dell	dellə	delle	dei	degli	dellɜ
alla	al	allo	all	allə	alle	ai	agli	allɜ
dalla	dal	dallo	dall	dallə	dalle	dai	dagli	dallɜ
nella	nel	nello	nell	nellə	nelle	nei	negli	nellɜ
sulla	sul	sullo	sull	sullə	sulle	sui	sugli	sullɜ

Table 15: Gendered and non-binary forms for articulated prepositions.