Reference editing guidelines

The **task** consists in editing a set of Italian references originally taken from GATE (Rarrick et al., 2023) so as to replace every linguistic expression of gender with a tag from a tagset covering gendered grammar. The tagset is given below, in Table 1 .

The Sheet file for this task includes several columns, three of which contain Italian reference translations for corresponding English source sentences. Column F (**REF-M**) contains masculine-only references, column G (**REF-F**) contains feminine-only references. These two references are perfectly identical, except for the words which express the gender of human beings. Finally, column H (**REF-TAGGED**) contains a copy of the masculine references, which are the ones that must be edited.

#	SOURCE	REF-M	REF-F	REF-TAGGED	CHECKREF
1	This year I got sick twice already.	Quest'anno mi sono ammalato già due volte.	Quest'anno mi sono ammalata già due volte.	Quest'anno mi sono ammalat<ENDS> già due volte.	Done
2	The student's essay was interesting and thoughtful, so the student got the highest grade possible.	Il saggio dello studente era interessante e riflessivo, quindi lo studente ha ricevuto il voto più alto possibile.	Il saggio della studentessa era interessante e riflessivo, quindi la studentessa ha ricevuto il voto più alto possibile.	Il saggio dello studente era interessante e riflessivo, quindi lo studente ha ricevuto il voto più alto possibile.	In progress
3	Compulsive gamblers are experts in self-deception.	I giocatori compulsivi sono esperti di autoinganno.	Le giocatrici compulsive sono esperte di autoinganno.	I giocatori compulsivi sono esperti di autoinganno.	In progress

The Sheet also includes a **CHECKREF** column which checks if the REF-M and the REF-TAGGED for that entry are different or identical, in order to visually keep track of progress.

What to edit

Within the REF-TAGGED reference, **each** word expressing gender and **only** the words expressing gender must be edited. More precisely, and as a guiding principle, only the words which differ in the REF-M and the REF-F should be edited in the REF-TAGGED.
Consider the example above:

REF-M	Quest'anno mi sono **ammalato** già due volte.
REF-F	Quest'anno mi sono **ammalata** già due volte.

The references are identical except for the word *ammalato*/*ammalata* (*sick*), which expresses the gender of the speaker. Therefore, that word must be edited in the REF-TAGGED.
Each REF-TAGGED includes at least one gendered word to be edited.
Epicene and gender-neutral terms, such as *ospite*/*ospiti* (*host*/*hosts*) and *persona*/*persone*, (*person*/*people*) must be ignored.

NOTE: There are a couple of specifications to keep in mind for this task, which are due to either design choices in the original data source or the scope of this test set:

- While the elided form of the definite articles *lo* and *la*, (*l'*) does not require any annotation as it does not allow to infer the gender of the referent – thus, only the subsequent noun/adjective must be edited – the elided form of the indefinite article *un*/*un'* must be appropriately tagged.
- While the term *studente*/*studenti* is gender-neutral in theory, it is regarded as masculine in the original dataset upon which this annotation is based, in opposition to the feminine forms *studentessa*/*studentesse*.

How to edit

Editing must be confined to the gendered terms identified according to the explanation above. Once a gendered word is identified, it should be edited in one of two ways, depending on the class of the word (open or closed class).

Open-class words editing

If it is an open-class word (a noun, a verb, or an adjective) only the gender morpheme must be edited, leaving the rest of the word – the root – unaltered. Consider the following example:

REF-M	Non sono mai **stat<u>o</u>** così bene.
REF-F	Non sono mai **stat<u>a</u>** così bene.
REF-TAGGED*	Non sono mai stato così bene.

*The asterisk indicates that the REF-TAGGED has not been tagged yet, thus it is incorrect.

In this case, only the word *stato* must be edited in the REF-TAGGED. As it is a verb, thus an open-class word, only the gender morpheme must be edited. Therefore, the root *stat-* will be left unaltered, and the morpheme *-o* will be replaced with the appropriate tag, in this case <ENDS> (see the Tagset section below):

REF-TAGGED	Non sono mai **stat<ENDS>** così bene.

In some instances the root of the word is less apparent, as in the pairs *senatore*/*senatrice* and *professore*/*professoressa*, where the masculine and the feminine forms do not differ for just one vowel but for a longer sequence. In these instances, the masculine form must be used to derive the root by removing the final morpheme, which can be easily identified by comparing the singular and the plural forms of the word. Consider the following example:

REF-M	Sono **professore** per vocazione.
REF-F	Sono **professoressa** per vocazione.

| MASC | professore | → | MASC-SING | professor**e** |
| FEM | professore**ssa** | | MASC-PLUR | professor**i** |

The comparison between the masculine and the feminine versions here is not useful to identify the part of the word that must be edited. Instead, by comparing the differences between the singular and the plural forms the final morpheme becomes apparent, thus indicating that the word must be edited as professor`<ENDS>` (see the Tagset section below).

Closed-class words editing

The closed-class includes articles, prepositions, possessive adjectives, demonstrative adjectives, and pronouns. Words belonging to this class must be replaced entirely with the appropriate tags. Consider the following example:

REF-M	Oggi devo incontrare **il mio** responsabile.
REF-F	Oggi devo incontrare **la mia** responsabile.
REF-TAGGED*	Oggi devo incontrare il mio responsabile.

Here, we have two closed class words to be edited: the article *il* and the possessive adjective *mio*. In the REF-TAGGED they must simply be replaced with the appropriate tags, in this case `<DARTS>` and `<POSS1S>` respectively (see the Tagset section below):

| REF-TAGGED | Oggi devo incontrare **<DARTS> <POSS1S>** responsabile. |

Tagset

Tags are uninterrupted sequences of characters starting with a < and ending with a >, e.g., `<DARTS>`. The tagset below attempts to represent all possible gendered terms in the grammar, but not all of the tags are necessarily represented in the testset. Only use the ones required.
The following table describes the whole tagset, providing an explanation of each tag and a mapping with respective masculine and feminine counterparts. Only the first two tags (`<ENDS>` and `<ENDP>`) should be used for open-class words, while the rest should be used for closed-class words.

TAG	Explanation	Masc	Fem
<ENDS>	Word end, singular	o, e, tore	a, essa, trice
<ENDP>	Word end, plural	i, tori	e, esse, trici
<DARTS>	Definite article, singular	il, lo, l'	la, l'
<DARTP>	Definite article, plural	i, gli	le
<IART>	Indefinite article	uno, un	una, un'
<PARTS>	Partitive article, singular	del, dello, dell'	della, dell'

<PARTP>	Partitive article, plural	dei, degli	delle
<PRON>	3rd person singular pronoun	lui	lei
<PREPdiS>	Articulated preposition: di + singular article	del, dello, dell'	della, dell'
<PREPdiP>	Articulated preposition: di + plural article	dei, degli	delle
<PREPaS>	Articulated preposition: a + singular article	al, allo, all'	alla, all'
<PREPaP>	Articulated preposition: a + plural article	agli, ai	alle
<PREPdaS>	Articulated preposition: da + singular article	dal, dallo, dall'	dalla, dall'
<PREPdaP>	Articulated preposition: da + plural article	dagli	dalle
<PREPinS>	Articulated preposition: in + singular article	nel, nello, nell'	nella, nell'
<PREPinP>	Articulated preposition: in + plural article	negli	nelle
<PREPsuS>	Articulated preposition: su + singular article	sul, sullo, sull'	sulla, sull'
<PREPsuP>	Articulated preposition: su + plural article	sugli	sulle
<DADJquelS>	Demonstrative adjective/pronoun (far) singular	quel, quello, quell'	quella, quell'
<DADJquelP>	Demonstrative adjective/pronoun (far) plural	quegli	quelle
<DADJquestS>	Demonstrative adjective/pronoun (near) singular	questo, quest'	questa, quest'
<DADJquestP>	Demonstrative adjective/pronoun (near) plural	questi	queste
<POSS1S>	Possessive adjective/pronoun, 1st p. s., singular	mio	mia
<POSS1P>	Possessive adjective/pronoun, 1st p. s., plural	miei	mie
<POSS2S>	Possessive adjective/pronoun, 2nd p. s., singular	tuo	tua
<POSS2P>	Possessive adjective/pronoun, 2nd p. s., plural	tuoi	tue
<POSS3S>	Possessive adjective/pronoun, 3rd p. s., singular	suo	sua
<POSS3P>	Possessive adjective/pronoun, 3rd p. s., plural	suoi	sue
<POSS4S>	Possessive adjective/pronoun, 1st p. p., singular	nostro	nostra
<POSS4P>	Possessive adjective/pronoun, 1st p. p., plural	nostri	nostre
<POSS5S>	Possessive adjective/pronoun, 2nd p. p., singular	vostro	vostra
<POSS5P>	Possessive adjective/pronoun, 2nd p. p., plural	vostri	vostre
<PRONDOBJS>	Direct object pronoun, singular	lo	la
<PRONDOBJP>	Direct object pronoun, plural	li	le
<PRONIOBJS>	Indirect object pronoun, singular	gli	le
<PRONIOBJP>	Indirect object pronoun, plural	gli	gli

Table 1: the tagset to be used to edit the Italian references. The tags are listed in the leftmost column and described in the second one. The two rightmost columns include the masculine and feminine words or segments that must be replaced with each tag.