Graphemic Methods for Gender-Neutral Writing

Yannis Haralambous & Joseph Dichy

Abstract. In this paper we present a model and a classification of graphemic gender-neutral writing methods, we explore current practices in French, German, Greek, Italian, Portuguese and Spanish languages, and we investigate interactions between gender-neutral writing forms and regular expressions.

1. Introduction: The General Issue of, and behind, Gender-Neutral Writing

The issue behind gender-neutral writing is that of the representation of inter-gender relations carried by languages. What is at stake is the representation of equality, or not, between genders. The issue is also referred to as "inclusive writing," which apparently refers to human rights, but does not cover all cases, i.e., lesbian, gay, bisexual and trans-gender persons. The term "Gender-Neutral Writing" is, in fact, both clearer and more inclusive.

Generally speaking, languages are conservative, if not archaic. A significant example is that of the idea of time, which is traditionally represented as a dot sliding along a straight line in a continuous movement. This has been a philosophical image of time since ancient Greek philosophers and throughout the Middle Ages both in Arabic and European philosophy, but can no longer be considered as a valid representation after 20th century existentialist philosophers and Heidegger's *Sein und Zeit*.

Yannis Haralambous 0000-0003-1443-6115
IMT Atlantique & LabSTICC UMR CNRS 6285
Brest, France
yannis.haralambous@imt-atlantique.fr

Joseph Dichy
Professor of Arabic linguistics
Lyon, France
joseph.dichy@yahoo.fr

Y. Haralambous (Ed.), *Graphemics in the 21st Century. Brest, June 13-15, 2018. Proceedings*
Grapholinguistics and Its Applications (ISSN: 2534-5192), Vol. 1.
Fluxus Editions, Brest, 2019, p. 41–89. https://doi.org/10.36824/2018-graf-hara2
ISBN: 978-2-9570549-0-9, e-ISBN: 978-2-9570549-1-6

FIGURE 1. Use of gender-neutral writing on the Web page of IMT Atlantique as a means to attract engineering students of both genders ("Bring out the engineer$_{gn}$ in you!")

This is even more true considering Einstein's relativity or any representation in modern physics. Nevertheless, the dot-moving-along-a-line representation remains efficient in the study of time and aspect in natural languages, because language structures reflect the naive or archaic view of human language users.

When it comes to gender, needless to say, the representation of inter-gender relations in linguistic lexical and grammatical structures as well as in discourse also remains archaic. Inclusive writing proposes to change the representation of genders in a way that puts forward equality between women and men. It nevertheless raises many questions, some of which are of a social and/or ideological nature, including educational aspects, while others are purely linguistic and technical.

2. Language and Ideology

2.1. Some Educational Issues

Inclusive writing is expected by many—albeit not by all—to have an ideological impact on gender equality which it endeavors to represent visually. It is also likely to have a positive educational effect on children and teenagers in schools.

The question remains of the oral utterance of graphemic gender-neutral writing. Oral strategies need to be devised, for instance: <les ambassadeur·rice·s> would become orally "les ambassadeurs et les ambassadrices" (with words in alphabetical order, according to the principles presented in Abily et al. 2016).

Let us take a parallel example in Arabic, where "pioneers" translates into *ruwwâd* رواد in the masculine form (resulting from the application of the plural pattern *fuʿâl* فعّاد to the singular *râʾid* رائد), and *râʾidât* رائدات in the feminine (by adding the suffix *–ât* ات to the singular). The difference between these two forms is related to Semitic morphology, which resorts, in Arabic, to modifications in patterns and in suffixes, and sometimes in both. The result—which is borrowed from a conference in the United Arab Emirates—was the translation of "pioneers of innovation" into: *râʾidât wa-ruwwâd al-ibtikâr* رائدات ورواد الإبتكار, where two Arabic words translate the English epicene "pioneers".

Generally speaking, the matter raised by gender-neutral writing is that of the relations between language and things. Some people, for example, will not use the word "cancer" because they are unconsciously in

the grip of the idea that if you say, "So&so has cancer," then So&so is closer to getting it. Coming back to our question, the epicene masculine appears as representing humanity as essentially made of male people. Eleanor Roosevelt (Gaer, 2009) just after the end of World War II imposed in the UN Charter the term "Human Rights," instead of "Man's Rights" ("les droits de l'homme" in French). The underlying idea was, and still is, that humanity is not by essence made of men.

On the other hand, from a purely philosophical and linguistic point of view, gender-neutral writing is a regression in the arbitrary relation between words (or signs) and things (the signs refer to), as if one needed a law in order to enforce linguistic equality between genders, while such an equality could be represented in languages using epicene words.

Nevertheless, considering the many steps which are still required for humanity to accept full equality between genders, such a philosophical or linguistic regression nowadays emerges as a social necessity. The issue is educational: teaching children at school that genders are equal is a question of quite some momentum.

2.2. Linguistic and Graphemic Aspects: Written Utterances That Cannot Be Vocalized

The main question in the linguistics of writing is that graphemic gender-neutral writing occurrences cannot be orally uttered. Let us take two examples of non-oralisable writing in French:

- Next to the Gare du Nord in Paris, a Moroccan gentleman, named Mr Binebine, used to sell, repair, and install taps (<robinet> in French). He had the idea of putting at the front of his store <RoBINEterie BINE> using capital letters to refer to his name and include it graphically.
- In a French university, the name of the "Faculty of Languages" (<Faculté des Langues>) was turned into: <faculté des ■ang■es>, where <langues> is replaced by <anges>, angels. The use of the black square as a meta-glyph is probably inspired by techniques of censorship or of marking unknown glyphs.

These two examples are not directly related to gender-neutral writing, but they recall that written utterances are not directly related to their oral realization. An advertisement for unisex clothes could be seen in France during the first months of 2018 for a trademark, the name of which was represented as LIU·JO,[1] thus featuring the middle dot of graphemic gender-neutral writing which we will present below. The

1. http://www.liujo.com/fr/

symbolic meaning of that point here between the names LIU and JO directly refers to the fact that both names can be either masculine or feminine, and that the trademark is "inclusively unisex".

3. A Formal Model for Graphemic Gender-Neutral Writing Methods

There are many gender-neutral writing methods: use of both genders ("ladies and gentlemen"), use of gender-neutral words ("people," "person," "individual," etc.), use of gender-neutral pronouns, like, in English, the Spivak pronouns: <e>, , <eir>, <eirs>, <emself> (`https://en.wikipedia.org/wiki/Spivak_pronoun`). In this paper we will deal with gender-neutral writing involving special graphemes and hence not representable in speech: French *écriture inclusive* <administrateur·rice>, German *binnen-I* <KollegInen>, Spanish *arroba* <trabajador@s>, Greek slashed suffixes <φοιτητές/τριες>, etc. We call these methods *graphemic gender-neutral writing methods*.

In this section we propose a formal model of graphemic gender-neutral writing methods which encompasses French, German, Greek, Italian, Portuguese and Spanish approaches.

NOTATION 1. — *Let <> denote graphemes. Let w be a two-gender singular-number word in French, German (nouns only), Greek, Italian, Portuguese or Spanish, $w_{♂,s}$ its singular-number masculine, $w_{♀,s}$ its singular-number feminine, $w_{♂,p}$ its plural-number masculine and $w_{♀,p}$ its plural-number feminine form. We use the following notation:*

- *let $C(w)$ be the common phonemic prefix of $w_♂$ and $w_♀$, i.e., if $g_1, g_2, \ldots, g_{i_m}$ are graphemes (or digraphs), g_i representing the phonemes of $w_♂$ and $g'_1, g'_2, \ldots, g'_{i_f}$ are graphemes (or digraphs) representing the phonemes of $w_♀$, then $C(w) := g_1, g_2, \ldots, g_{i_c}$ such that $g_j = g'_j$ for all $1 \leq j \leq i_c$ and $i_c \leq \min(i_m, i_f)$;*[2]
- *in some cases, $C(w)$ differs slightly between the masculine and the feminine version, we will denote these by $C_♂(w)$ and $C_♀(w)$;*
- *let $\varphi_♂$ and $\varphi_♀$ be transformations of grapheme chains;*
- *let* SEPG *be a special grapheme called* separator grapheme*;*
- *let <+> denote the grapheme string concatenator;*
- *let $SS_♂(w)$ and $SS_♀(w)$ be the gender-specific suffixes of $w_{♂,s}$ and $w_{♀,s}$, i.e., $w_{♂,s} = C(w_s) + S_{♂,s}(w)$ and $w_{♀,s} = C(w_s) + S_{♀,s}(w)$. Similarly, let $SP_♂(w)$ and $SP_♀(w)$ be the gender-specific suffixes of $w_{♂,p}$ and $w_{♀,p}$.*

2. Note that i_c need not be maximal, as in Greek <φοιτητές/τριες> where the <τ>, albeit common to both suffixes <τές> and <τριες>, is not part of $C(w)$.

DEFINITION 1. – *With the notation above, we define the* gender-neutral singular-number form $GN_s(w)$ *of* w*, and the* gender-neutral plural-number form $GN_p(w)$ *of* w *as:*

$$GN_s(w) := C(w) + \varphi_{\sigma}(SS_{\sigma}(w)) + \text{SEPG} + \varphi_{\varphi}(SS_{\varphi}(w)),$$
$$GN_p(w) := C(w) + \varphi_{\sigma}(SP_{\sigma}(w)) + \text{SEPG} + \varphi_{\varphi}(SP_{\varphi}(w)).$$

Here is how this model can be applied to the six languages we are considering, to obtain the singular-number gender-neutral form of a word:

Language	$C(w)$	$SS_{\sigma}(w)$	$\varphi_{\sigma}(SS_{\sigma}(w))$	SEPG	$SS_{\varphi}(w)$	$\varphi_{\varphi}(SS_{\varphi}(w))$	$GN_s(w)$
French	act	eur	eur	·	rice	rice	acteur·rice
German	Student				in	In	StudentIn
Greek	μαθη	τής	τής	/	τρια	τρια	μαθητής/τρια
Italian	ragazz	o			a	@	ragazz@
Portuguese	alem	ão	ão	/	ã	ã	alemão/ã
Spanish	abogad	o			a	@	abogad@

Similarly, here is how the plural-number form is obtained:

Language	$C(w)$	$SP_{\sigma}(w)$	$\varphi_{\sigma}(SP_{\sigma}(w))$	SEPG	$SP_{\varphi}(w)$	$\varphi_{\varphi}(SP_{\varphi}(w))$	$GN_p(w)$
French	act	eurs	eur	·	rices	rice·s	acteur·rice·s
German	Student				innen	Innen	StudentInnen
Greek	μαθητ	ές	ές	/	τριες	τριες	μαθητές/τριες
Italian	ragazz	i			e	@	ragazz@
Portuguese	alem	ães	ães	/	ãs	ãs	alemães/ãs
Spanish	abogad	os			as	@s	abogad@s

3.1. Graphemic Approaches to Gender-Neutral Writing

There are three main approaches to graphemic gender-neutral writing: the Single-Grapheme Replacement method SINGLE (Italian, Spanish), the Marked Feminine Suffix method MARK (German) and the Suffix-Join JOIN method (French, Greek, Portuguese).

3.1.1. SINGLE

In the *Single Grapheme Replacement* method (SINGLE), one or more gender-specific graphemes are replaced by a single, gender-neutral grapheme, which we call *gender replacement grapheme* (REPG) (e.g., in Spanish, <abogad@s> being the gender-neutral form of <abogados> and <abogadas>). The SINGLE method is used in Italian, Spanish and occasionally in Portuguese.

The cognitive load of the SINGLE method depends on the variety of graphemes represented by the replacement grapheme: if it always represents the same pair of graphemes (for the feminine and the masculine version of the word) then its decoding is easier than if it may need to represent various pairs of graphemes and the reader needs to choose among them to rebuild the original word.

3.1.2. MARK

In the *Marked Feminine Form* (MARK) method, the feminine form—and more generally the feminine grammatical role—is used. To denote gender neutrality, the first grapheme of the feminine suffix is marked, most often by case inversion (as in German <StudentInnen>, or <STUDENTiNNEN>, which are the gender-neutral forms of <Studenten> and <Studentinnen>), but possibly also by preceding it by an underscore (the *gender gap*) or by an asterisk (the *gender star*). The MARK method is used in German.

3.1.3. JOIN

In the *Suffix Join* (JOIN) method, the gender-specific suffixes are joined after the stem and are separated by a specific grapheme, called *gender separator grapheme* SEPG (as in French <étudiant·e>, which is the gender-neutral form of <étudiant> and <étudiante> with a middle dot as separator grapheme, or in Greek <νέος/α>, which is the the gender-neutral form of <νέος> and <νέα> with a slash as separator grapheme).

The JOIN approach is used in French (*écriture inclusive*), Greek and Portuguese.

The cognitive load of the JOIN method depends mainly on the size of the first suffix which has to be mentally deleted by the reader in order to obtain the version using the second suffix. Therefore, to evaluate JOIN methods for different languages, we introduce the following notion:

DEFINITION 2. – *In a JOIN method, we call* backtrack *(BT) the length of the first suffix.*

For example, the backtrack of <administrateur·rice> is 3 since the suffix <eur> of length 3 has to be removed in order to obtain the feminine version <administratrice>. Determination of common prefix and backtrack is done separately for the singular and for the plural number. For example, in the singular of the Portuguese word <cantonês>/<cantonesa>, the common prefix is <canton> (BT=2), while its <cantoneses>/<cantonesas> will have a common prefix <cantones> (and therefore again BT=2).

3.2. Gender Symmetry and Asymmetry

Definition 3. – *Let w be a word. We call a graphemic gender-neutral form $GN(w)$* gender symmetric *if, grammatically and visually, $GN(w)$ is at equal distance from $w_{\male}$ and $w_{\female}$.*

If $GN(w)$ is grammatically or visually closer to $w_{\female}$, we call it ♀-privileging.

If it is grammatically or visually closer to $w_{\male}$, we call it ♂-privileging.

Single methods are mostly symmetric, for example, the Spanish <abogad@s> is *symmetric* because it is equally close to <abogados> and to <abogadas>, where <close> can be either the Levenshtein distance (both strings can be obtained by a single-character substitution) or the visual resemblance of <@> with <o> and <a> (indeed the grapheme <@> has the shape of an <a> contained in a <o>).

As we will see in § 7, the German Mark method is globally ♀-privileging since it is basically the feminine form. In the case of <StudentInnen> is visually very close to the feminine <Studentinnen>. The asymmetry is even stronger for words with umlauted feminine form: in these cases $C(w)$ is gender-specific and therefore the choice of using its feminine form brings $GN(w)$ closer to $w_{\female}$ than to $w_{\male}$, e.g., if $w_{\female}$ is <Ärztin> (with umlauted <Ä>) and $w_{\male}$ is <Arzt> (no umlaut), then $GN(w)$ is <ÄrztIn>, which is visually closer to <Ärztin> than to <Arzt>.

Join methods lack symmetry because an order has to be chosen between masculine and feminine suffix. Indeed, as the linear chaining of graphemes (mostly) reflects their temporal succession, Join methods can *never* be symmetric: one of the two suffixes has to be written first and ipso facto becomes privileged. For example, French *écriture inclusive*, when using the ♂♀ order, is ♂-privileging: <étudiant·e>, <administrateur·rice>.

We discuss the order of suffixes for the French Join method in 8.3 and for the Greek Join method in 10.1.

4. Hypotheses for Graphemic Gender-Neutral Writing Methods

4.1. Hypotheses for the Single Method

In order to apply the Single method, we need the following hypothesis to be valid:

Hypothesis 1 (Strong Single Hypothesis). – *Both in the singular and in the plural number, the masculine and feminine versions of a given two-gender word differ by a single grapheme.*

To obtain a gender-neutral version of a given word, it suffices to replace that grapheme by a specific gender-neutral and easily identifiable

grapheme, the *replacement grapheme* REPG. We will investigate for each language: (a) the percentage of words (nouns and adjectives) that satisfy the Strong SINGLE Hypothesis, and (b) whether the proposed replacement grapheme is compatible with the hypothesis.

In some cases there are additional differences between the masculine and feminine versions of a word. For example, a vowel may be accented in one case and not in the other, or an additional letter may appear in front of the replacement grapheme in one case and not in the other, or the replacement grapheme may stand for an empty grapheme in one case and not in the other. For these reasons we state a weaker version of the hypothesis:

HYPOTHESIS 2 (Weak SINGLE Hypothesis). – *Both in the singular and in the plural number, the masculine and feminine versions of a given two-gender word differ by a single grapheme (which may be missing or may be preceded by some other grapheme in one of the two genders) and potentially by the presence or absence of an accent on a grapheme of the common prefix.*

This covers cases such as the Italian masculine <arcaici>, the feminine of which is <arcaiche> (a letter <h> is added), the Italian <figli>, the feminine of which is <figlie> (the replacement grapheme <@> in <figli@> stands for an empty grapheme in the masculine version), and the Spanish <mocetón>, the feminine form of which is <mocetona> (without the acute accent).

Gender-neutral words that satisfy the weak SINGLE hypothesis and not the strong one are asymmetric: in the three examples above, <arcaic@> is ♂-privileging (since the absence of the <h> brings the gender-neutral form closer to the masculine one) while <arcaich@> is ♀-privileging; <figl@> is ♂-privileging, while <figli@> is ♀-privileging; <mocetón@> is ♂-privileging, while <moceton@> is ♀-privileging.

4.2. Hypotheses for the MARK Method

In order to apply the MARK method, we need the following hypothesis to be valid:

HYPOTHESIS 3 (Strong MARK Hypothesis). – *The singular feminine form of a word (noun or adjective) is equal to the singular masculine form followed by a suffix, which is the same for all words of the language. The plural feminine form of a word (noun or adjective) is equal to the singular masculine form followed by a suffix, which is the same for all words of the language.*

By marking the first grapheme of the suffix (either by case inversion, or by preceding it by a <_> or a <*> grapheme), we obtain a gender-neutral version of the word. This method relies on the fact that

all nouns of the language share the *same* suffix, otherwise it becomes difficult for the reader to make the connection between marked grapheme and gender-neutral intention. We will investigate whether this is the case for German.

We also state a weaker version of the hypothesis:

Hypothesis 4 (Weak Mark Hypothesis). – *The singular feminine form of a word (noun or adjective) is equal to the singular masculine form (after removing 0, 1 or 2 graphemes and possibly adding an umlaut) followed by a suffix, which is the same for all words of the language. The plural feminine form of a word (noun or adjective) is equal to the singular masculine form (after removing 0, 1 or 2 graphemes and possibly adding an umlaut) followed by a suffix, which is the same for all words of the language.*

This covers cases such as <Beamter>, the feminine version of which is not <*Beamterin> but <Beamtin> (two graphemes have to be removed from <Beamter> before adding the <in> suffix) or such as <Jude>, the feminine version of which is umlauted: <Jüdin>.

Mark gender-neutral forms are, by definition, asymmetric since they are visually closer to the feminine form, and hence are ♀-privileging. In the case of umlauted stems, this property is even stronger: <JüdIn> is closer to the feminine form <Jüdin> than the erroneous *<JudIn>, in which the stem has not been umlauted.

4.3. Hypotheses for the Join Method

Finally, in order to apply the Join method, we need the following hypothesis to be valid:

Hypothesis 5 (Strong Join Hypothesis). – *Whether in singular or in plural number, the masculine and feminine forms of a word must have a common nonempty stem, to which a (possibly empty) suffix has to be added in order to obtain the masculine form, and a different suffix has to be added in order to obtain the feminine form.*

We obtain the gender-neutral form by writing the common stem followed by either the masculine or the feminine suffix, then a separator grapheme sepg and, finally, the other suffix. This hypothesis makes no assumption on the order of suffixes. We call the length of the first suffix *backtrack*.

If we don't require nonemptyness of the common prefix, then any two words can be combined to form a gender-neutral form, even if they have nothing in common, such as <femme·homme> or <fille·garçon>, so the hypothesis is necessarily *true for all words*. Nevertheless, to keep the cognitive load as low as possible, there are cases where we may ignore slight differences in stems. For example, in the case of Greek words, stems may differ only by accent position, as in the masculine <φοιτητῆ> (accented on the ultima) and the feminine <φοιτήτριας> (accented on

the penult); one can disregard the diacritic and write <φοιτητῆ/τριας> instead of the longer <φοιτητῆ/τήτριας>. To cover this case, we state a weaker version of the Join hypothesis:

Hypothesis 6 (Weak Join Hypothesis). – *Whether in singular or in plural number, the masculine and feminine forms of a word must have a common (modulo accent position) nonempty stem, to which a (possibly empty) suffix has to be added in order to obtain the masculine form, and a different suffix has to be added, in order to obtain the feminine form.*

This makes it possible for the writer to write accent-independent gender-neutral forms <βάτραχος/ίνα> (♂♀ order) and <βατραχίνα/ος> (♀♂ order) with backtracks 2 and 3, instead of the absurd *<βάτραχος/ατραχίνα>, *<βατραχίνα/άτραχος>, which would have backtracks 7 and 8 (!!).

Here again, weakness of the hypothesis can increase the asymmetry of the gender-neutral form and cognitive load. This is not the case in our first example <βάτραχος/ίνα> since accents of both forms appear in the gender-neutral form and guide the reader into reconstructing the gender-specific forms. It is, however, strongly the case in the second example <βατραχίνα/ος> since the reader has to reconstruct the masculine form by mentally repositioning the accent to the penult (<βάτραχος>).

Let un now consider graphemic gender-neutral writing methods for various languages: Italian (§ 5), Spanish (§ 6), German (§ 7), French (§ 8), Portuguese (§ 9), and Greek (§ 10).

5. The Italian Single Method

The Italian Single approach consists in replacing the final vowel of Italian nouns and adjectives by a replacement grapheme, which can be <@> or <*>, e.g., <ragazz@ italian@> or <ragazz* italian*>, instead of <ragazzi italiani e ragazze italiane>.

5.1. History

The feminist publication (Not One Less, 2017) uses the <@> sign (called *chiocciola*) as a gender-neutral graphemic replacement of -o/-a (singular) or -i/-e (plural):

> In questo Piano abbiamo scelto di svelare la non neutralità del maschile utilizzando non solo il femminile, ma anche la @ per segnalare l'irriducibilità e la molteplicità delle nostre differenze.[3]

In the 57-page long (ibid.) booklet the gender-neutral grapheme <@> is used 50 times for 23 different words, mostly for the words <tutt@> ("all$_{gn}$"[4]) and <ognun@> ("nobody$_{gn}$"). When, on p. 12, an article has is written in gender-neutral form, a slash-based JOIN method is applied instead: <delle/degli altr@> ("of the$_{gn}$ others$_{gn}$").

In 2012, in a report financed by the Region of Tuscany and supported by the *Accademia della Crusca*, Cecilia Robustelli (2012) mentions the asterisk <*> as a gender-neutral grapheme, but discourages its use:

> L'uso di forme abbreviate attraverso altri espedienti grafici, come per esempio l'inserimento dell'asterisco al posto della desinenza per indicare che si intende sia la forma maschile sia quella femminile, es. ragazz* anziché ragazzo/ragazza o ragazzo/a, è da evitare perché può ostacolare la lettura e la comprensione del testo.[5]

Notice that the JOIN slash-based approach <ragazzo/a> is also mentioned.

We have investigated the validity of the strong and weak SINGLE hypotheses for the Italian language. For this we have extracted from the Italian Wiktionary (version of July 20th, 2019) the declension tables of 12,379 Italian nouns, adjectives and participles. Here are the results, according to their compatibility with the two versions of the hypothesis:

Words Compatible with the Strong SINGLE Hypothesis

We have found the following word classes validating the strong SINGLE hypothesis:

---

3. "In this plan we have chosen to reveal the non-neutrality of the masculine form by using not only the feminine form, but also the @ sign to signal the irreducibility and multiplicity of our differences."

4. In this text we will mark gender or gender neutrality in English translations as follows: (a) if the word is of feminine or masculine gender in the source language, its translation will carry the subscript "$_{♀}$" or "$_{♂}$," resp., e.g., we translate <pronto> by "ready$_{♂}$"; (b) if the word is gender-neutral or epicene in the source language, its translations will carry the subscript "$_{gn}$" or "$_{epi}$" respectively, e.g., we translate <pront@> by "ready$_{gn}$"; (c) and if the word is a gender-neutral personal pronoun, to translate it into English we will use Spivak personal pronouns, e.g., we translate <il/elle parle> by "e talks".

5. "The use of abbreviated forms through other graphic expedients, as for example the insertion of an asterisk instead of the suffix to indicate that both masculine and feminine forms are meant, e.g., *ragazz** instead of *ragazzo/ragazza* or *ragazzo/a*, is to be avoided because it can hinder reading and understanding of the text."

Class	Suffixes	#	Typical example
1	o/a, i/e	9,138	tutto/tutta, tutti/tutte
2	a/a, i/e	419	ciclista/ciclista, ciclisti/cicliste
3	e/e, i/i	175	inutile/inutile, inutili/inutili
4	o/a, hi/he	157	stucco/stucca, stucchi/stucche
5	e/a, i/e	132	cantoniere/cantoniera, cantonieri/cantoniere
	Total	10,021	

These classes cover 81% of the total amount of words. Class 2 is epicene in the singular number, and Class 3 is epicene in both numbers.

Words Compatible with the Weak SINGLE Hypothesis

We found the following word classes validating the weak but not the strong hypothesis:

Class	Suffixes	#	Typical example
6	o/a, i/he	1,515	arcaico/arcaica, arcaici/arcaiche
7	o/a, ∅/e	353	figlio/figlia, figli/figlie
	Total	1,868	

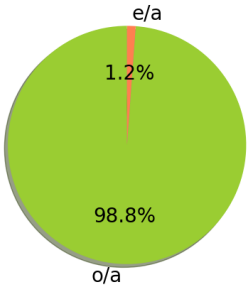
While "strong" Classes 1–5 are totally symmetric (the form <tutt@> can equally well represent <tutti> or <tutte>) , this not the case of Classes 6 and 7.

In the plural of Class 6, the writer has to choose between <arcaic@> and <arcaich@>. The former is ♂-privileging (since the absence of <h> makes <arcaic@> closer to <arcaici> than to <arcaiche>) and the latter ♀-privileging.

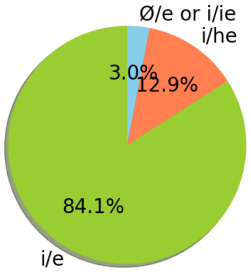
In Class 7, the choice is between <figli@> and <figl@>, the first being ♀-privileging (since it is more plausible that <@> stands for letter <e> than for a missing letter as in the masculine <figli>) and the second ♂-privileging. Ironically, the feminist document (Not One Less, 2017) uses the second form, which is ♂-privileging.

Polysemy of the REPG

As we can see in the following diagrams, the semantics of the replacement grapheme are very stable in the singular, since in 98.8% of cases it represents the same pair of vowels o/a. In the plural, only in 84.1% of cases does it represent the pair i/e, while in 12.9% of cases, if we follow the ♂-privileging approach, an <h> appears (as in <arcaic@> representing <arcaici> and <arcaiche>).



REPG semantics distribution in the singular



REPG semantics distribution in the plural

Forms Incompatible with Both SINGLE *Hypotheses*
Among the 490 words we found that are incompatible with both versions of the SINGLE hypothesis (that is 4% of the total number of words), let us mention just one class:

Class	Suffixes	#	Typical example
8	ore/rice, ori/rici	422	traduttore/traduttrice, traduttori/traduttrici

Here the suffix differs by significantly more than one letter and hence the SINGLE method cannot be applied. The only possible solution would be to use a JOIN method with a backtrack of 3 letters, to obtain <traduttore/rice>, <traduttori/rici> which, of course, is ♂-privileging, since the masculine form comes first. This class represents 86% of the set of incompatible words and 3.4% of the total set of words.

5.1.1. Conclusion

The strong SINGLE hypothesis is valid for 81% of Italian nouns and adjectives, while the weak SINGLE hypothesis is valid for 96% of them. We can reasonably conclude that the SINGLE approach is appropriate for the Italian language.

There is nevertheless a caveat: the replacement grapheme *represents different graphemes in the singular and the plural*, so that the reader must collect information from the context to identify the number of each gender-neutral form, in order to be able to decode it.

6. The Spanish SINGLE Method

The Spanish SINGLE approach involves replacing the vowel of the ultima of Spanish nouns and adjectives by a specific replacement grapheme

which can be <@> (*arroba*), or <*>, or <e>, or <x>: <niñ@s español@s>, or <niñxs españolxs>, or <nin*s español*s>, or <niñes españoles> instead of <niños españoles y niñas españolas>.

6.1. History

In a book called "Sexism and language" (García Meseguer, 1976)[6], García Meseguer suggests using <e> as replacement grapheme:

> Así, cuando une se dirija a un grupo en una conferencia, en una carta circular, etc., podrá comenzar diciendo *querides amigues*. *Les trabajadores* podrán escribir en sus pancartas reivindicativas *estamos hartes de ser explotades*. *Les políticques* podrán llamar *compañeres* a sus *partidaries*. *Les progenitores* podrán educar a sus *hijes* más fácilmente en forma no sexista. En los periódicos, los anuncios por palabras solicitarán *une cocinere*, *une abogade* o *une secretarie*.[7]

This proposal was followed recently by various politicians: on June 13, 2018, the Argentinian parlamentarian Marcos Cleri started a speech by <Buenas tardes a todes> ("good evening to everybody$_{gn}$") and used the <e> replacement grapheme in the entire speech. On June 25 of the same year, the former prime minister of Chile Michelle Bachelet wrote in a tweet: <los miles de chiquilles que hoy estudian con gratuidad en Chile>[8] In 2018 the National University of La Plata started a television program called *Todes* ("Everyone$_{gn}$"), and in April 2019 this university organized a Conference on Inclusive Language[9].

As for other graphemes than <e>, in 2009 the collective publication *Interdicciones, escrituras de la intersexualidad en castellano* ("Interdictions, writings of intersexuality in Spanish," Cabral 2009) uses the <*> grapheme in several texts.

In 2012 the University of Valencia published a "Guide for an egalitarian language" (Quilis Merín, Albelda Marco, and Josep Cuenca, 2012) where the usage of <@> is discouraged:

---

6. See also the detailed bibliography in `https://www.sexismoylenguaje.com/polemica-guias-para-un-uso-no-sexis`.

7. "Thus, when you join a group of people in a conference, in a collective letter, etc., you can start by saying *dear friends*$_{gn}$. *Workers*$_{gn}$ will be able to write in their claim placards "we are *tired*$_{gn}$ of being *exploited*$_{gn}$". *Politicians*$_{gn}$ may call their *supporters*$_{gn}$ *companions*$_{gn}$. *Parents*$_{gn}$ can educate their *children*$_{gn}$ more easily in a non-sexist way. In the newspapers, word ads will request a *cook*$_{gn}$, a *lawyer*$_{gn}$ or a *secretary*$_{gn}$."

8. "The thousands$_{gn}$ of kids$_{gn}$ who study today for free in Chile."

9. For further information on gender-neutral writing in Argentina, see (Patti, 2018).

> Evitar el desdoblamiento abreviado con barras y la arroba (@), a no ser que se trate de nombres propios de organismos, grupos o eventos que la hayan incorporado, o que se emplee como herramienta de diseño publicitario.[10]

In 2013, the Argentinian transvestite activist Lohana Berkins promoted the use of graphemes <@> or <x> in the Argentinian newspaper *Página|12* (Berkins, 2013).

In 2016 the goverment of Chile published a "Guide for a gender-inclusive language," in which the use of <@> is discouraged:

> El signo "@" no es lingüístico, rompe con las reglas gramaticales del idioma y es impronunciable por lo tanto su uso no es recomendable.[11]
> (National Council of Culture and Arts, 2016, p. 6)

6.2. Evaluation

In the following we investigate whether the SINGLE hypotheses are valid for the Spanish language. For this we have extracted from the Spanish Wiktionary (version of July 20th, 2019) the declension tables of 73,473 Spanish nouns and adjectives. Here are the following results according to their compatibility with the two versions of the hypothesis:

Invariant or Gender-Invariant Words
The following forms are either totally invariant, or epicene:

Class	Suffixes	#	Typical example
1	(invariant)	26,704	abrecoches
2	(gender inv.)	2,383	moralista, moralistas
	Total	29,837	

These classes cover 40.6% of the total number of words. We will consider that they validate the strong SINGLE hypothesis, since they need no special grapheme in the first place.

10. "Avoid segmentation with slashes or use of the *arroba* (@), with the exception of names of people, organisms, groups or events that have incorporated it, or are using them in advertising design."

11. "The sign '@' is not linguistic, breaks Spanish language grammar rules and is unpronounceable, therefore its use is not recommended."

Words Compatible with the Strong SINGLE Hypothesis
The feminine and masculine forms of the following word classes differ only in the final vowel, while adding an <s> for the plural:

Class	Suffixes	#	Typical example
3	o/a, os/as	37,434	abogado/abogada, abogados/abogadas
4	e/es, a/as	13	chilote/chilota, chilotes/chilotas
	Total	37,447	

These classes cover 51% of the total number of words.

Words Compatible with the Weak SINGLE Hypothesis
We have found the following word classes, in which the final vowel of the masculine singular form is either missing or different than in the other forms:

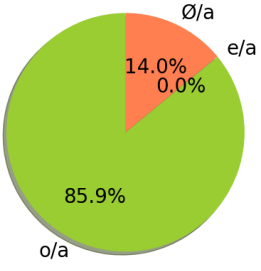
Class	Suffixes	#	Typical example
5	∅/a, es/as	3,369	trabajador/trabajadora, trabajadores/trabajadoras
6	ón/ona, ones/onas	2,616	mocetón/mocetona, mocetones/mocetonas
7	és/esa, eses/esas	95	montañés/montañesa, montañeses/montañesas
8	án/ana, anes/anas	6	alazán/alazana, alazanes/alazanas
9	ín/ina, ines/inas	6	chapín/chapina, chapines/chapinas
	Total	6,092	

These classes cover 8.3% of the total number of words.

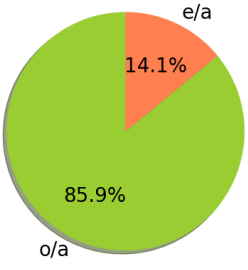
While "strong" Classes 1–4 are totally symmetric (the form <gauch@> represents <gaucho> and <gaucha> equally well and the plural form <gauch@s> represents <gauchos> or <gauchas> equally well) , this not the case of Classes 5–9. For example, in Class 5, the form <trabajador@> is ♀-privileging since it assumes the existence of a final vowel, which is only the case for the feminine <trabajadora>. In Class 7, in the singular case, the user has the choice between writing <montañés@> (which is ♂-privileging) or <montañes@> (which is ♀-privileging).

Words Incompatible with Both SINGLE *Hypotheses*
We found 8 nouns with a feminine in -riz: <acelerador>/<aceleratriz>, <actor>/<actriz>, <adorador>/<adoratriz>, <director>/<directriz>, <emperador>/<emperatriz>, <formador>/<formatriz>, <generador>/<generatriz>, <tutor>/<tutriz>,<tutriz>, for which only a JOIN method can be used: <actor/riz>, <actores/rizes>, etc., with a backtrack of 2 in the singular and 4 in the plural number.

Polysemy of the REPG
As we can see in the following diagrams, the semantics of the replacement grapheme are relatively stable in both the singular and the plural: in 85.9% of cases it represents the pair of vowels <o>/<a>. Contrary to Italian, for these 85.9% of cases, the replacement grapheme represents the same values <o>/<a> for singular and plural (since, in Spanish, a final <s> morpheme carries the plural information). So, in some sense, the feminine value of the replacement grapheme is always <a>, while the masculine is mostly <o> but can also be empty in the singular or <e> in the plural.



REPG semantics distribution in the singular



REPG semantics distribution in the plural

Forms Incompatible with the Use of the <e> Grapheme
Even though the grapheme <e> was historically the first proposed (as early as in 1976, see §6.1), there are many cases in which it cannot be used because the gender-specific suffix already contains an <e>:

- in Class 4, where in both the singular and the plural number, the masculine suffix contains <e>, and the feminine suffix does not;
- in Classes 5–9, where the masculine suffix of the plural number contains <e>, and the feminine suffix does not.

Using an <e> in these cases, which represent 8.3% of the total number of words, results in ambiguity between the gender-neutral and the masculine form: when writing <trabajadores>, is <e> the grapheme of

the masculine plural <trabajadores> or the special gender-neutral replacement grapheme? In the former case <trabajadores> refers to male workers only, while in the latter it refers to workers of both genders.

6.2.1. Conclusion

The strong SINGLE hypothesis is valid for 91.6% of Spanish nouns and adjectives, while the weak SINGLE hypothesis is valid for over 99% of them, when the gender-neutral replacement grapheme is <@>, <*> or <x>. In the case of the <e> grapheme, the weak SINGLE hypothesis cannot be applied, so the total ratio of words compatible with the SINGLE method is only 91.6%.

We conclude that the SINGLE approach is very well adapted to the Spanish language, when one of the gender-neutral replacement graphemes <@>, <*> or <x> is used, but is less efficient when the replacement grapheme <e> is used.

7. The German MARK Method

The German MARK method (called "binnen-I" or "binnenmajuskel") consists in marking the letter <I> of the feminine suffix of nouns, either by case-inverting it, e.g., <StudentInnen> (where <Studenten> is the masculine plural and <Studentinnen> the feminine plural), or by preceding it by a <_> grapheme or a <*> grapheme, as in <Student_innen> or <Student*innen>. Although it can also be applied to the singular, it is mostly used in the plural, as inclusive of both genders.

When preceded by articles and adjectives, the *feminine* grammatical gender is applied (as in <jede neue KollegIn>, "every$_{♀}$ new$_{♀}$ colleague$_{gn}$"), which makes this approach a ♀-privileging one, according to Kotthoff and Nübling (2018, p. 217):

> Wegen der Femininkongruenz wird das Femininum (bewusst) privilegiert.[12]

Indeed, the reader's eye will first recognize the feminine suffix, before (potentially) realizing that letter <I> is in upper case. As (Oestreich, 2009) puts it:

> Das Durchschnittsgehirn kennt nämlich keine Binnenmajuskel, also keinen Großbuchstaben inmitten eines Wortes und liest das I als kleinen

12. "The feminine feature is (consciously) privileged, because of gender agreement."

> Buchstaben. Bei PolitikerInnen liest es Politikerinnen – und fragt sich, wo da die Männer bleiben.[13]

As for the singular number, according to Kotthoff and Nübling (2018), adjectives, articles and pronoun dependencies of a gender-neutral noun should use the feminine form: <jede neue KollegIn> ("every$_{♀}$ new$_{♀}$ colleague$_{gn}$"). On the other hand, Damm et al. (2014, p. 16) proposes an alternative scheme, where the final grapheme of noun dependencies is also marked: <jedE neuE KollegIn>.

Whether we mark dependencies or not, there is an additional issue which is specific to the singular number, namely *declension*. Indeed, the genitive of the gender-neutral <die ProfessorIn> will be <der ProfessorIn>, which is quite different from the masculine genitive <des Professors>. The result is even more ♀-privileging, and it may be more interesting to write complete words <der Professorin oder des Professors> instead of writing <der ProfessorIn> and have the reader phonetically realize it as <der Professorin oder des Professors>.

7.1. History

According to (Schoenthal, 1998) the MARK method was used for the first time in a self-published 627-page book on pirate radios with instructions on how to build a radio station, published in 1981 (Busch, 1981). In 1983, the MARK method was first used by the Swiss weekly newspaper *Schweize Wochenzeitung* (WoZ) and the same year by the Berlin newspaper *Berliner Tageszeitung* (taz) (Kotthoff and Nübling, 2018, p. 218).

The MARK method is mentioned in the specific *Duden* on gender issues (Diewald and Steinhauer, 2017, p. 44):

> Diese Schreibung ist seit Anfang der 1980er-Jahre belegt und in bestimmten Kontexten sehr gebräuchlich. Allerdings sehen die offiziellen Rechtschreibregeln Binnengroßbuchstaben nicht vor; sie lehnen sie aber auch nicht explizit ab, denn die Binnengroßschreibung ist schlicht gar nicht Gegenstand des amtlichen Regelwerks.[14]

In §11 we discuss experimental versions of the MARK method.

---

13. "The average brain is not aware of *binnen*-letters, i.e., capital letters in the middle of a word, and reads the letter I as a lowercase letter. In the word *PolitikerInnen* (politicians$_{gn}$) it will read *Politikerinnen* (politicians$_{♀}$)—and will wonder why there are no men."

14. "This form of writing is documented since the early 80s and is very common is specific contexts. However, official spelling rules do not consider internal capital letters; but they do not explicitly prohibit them either, because internal capital writing is simply not an issue for official regulations."

7.2. Evaluation

To investigate the validity of the MARK hypotheses, we extracted data on 4,561 two-gender nouns from the German Wiktionary.

7.2.1. Strong MARK Hypothesis

We found 4,079 nouns (that is 89.4% of the total number of two-gender nouns) conforming with the strong MARK hypothesis. They can be subdivided into three classes:

Class	#	Typical examples
$GN(w) = C_{\sigma}(w) + \text{<In>}$	3,672	Student → StudentIn
$GN(w) = C_{\sigma}(w) - \text{last letter} + \text{<In>}$	368	Kollege → KollegIn
$GN(w) = C_{\sigma}(w) - \text{two last letters} + \text{<In>}$	39	Beamter → BeamtIn

If i is the number of letters to remove before adding the -<in> suffix, then $i = 0, 1, 2$ are the only possible cases in our corpus.

7.2.2. Weak MARK Hypothesis

Among the remaining 480 nouns, 136 validate only the weak MARK hypothesis, in the sense that the stem of the feminine noun is umlauted while the stem of the masculine is not. Again we have three classes:

Class	#	Typical examples
$GN(w) = \text{umlauted}(C_{\sigma}(w)) + \text{<In>}$	117	Arzt → Ärztin
$GN(w) = \text{umlauted}(C_{\sigma}(w)) - \text{last letter} + \text{<In>}$	18	Jude → Jüdin
$GN(w) = \text{umlauted}(C_{\sigma}(w)) - \text{two last letters} + \text{<In>}$	1	Tauber → Täubin

7.2.3. Cases Where the MARK Hypothesis Is Invalid

In 26 cases of Wiktionary two-gender nouns, the MARK method cannot be applied because the plural form is epicene: <Angeklagte>, <Elfe>, <Linke>, <Süße>, etc.

In 193 cases the incompatibility is of a lexical nature: the following gender-related antonymous pairs serve as bases for composite word creation:

Pair	#	Examples
Mann/Frau	54	Ehemann/Ehefrau, Fachmann/Fachfrau, etc.
Vater/Mutter	15	Stiefvater/Stiefmutter, Großvater/Großmutter, etc.
Sohn/Tochter	11	Pflegesohn/Pflegetochter, Enkelsohn/Enkeltochter, etc.
Bruder/Schwester	8	Knastbruder/Knastschwester, Vollbruder/Vollschwester, etc.
Junge/Mädchen	9	Bauernjunge/Bauernmädchen, Zeitungsjunge/Zeitungmädchen, etc.

In these cases non-graphemic solutions have to be sought.

Finally in about a hundred cases, words are of foreign origin and are feminized according to the rules of their original language: <Coiffeur>/<Coiffeuse>, <Cowboy>/<Cowgirl>, <Filipino>/<Filipina>, <Yogi>/<Yogini>, etc.

7.2.4. Conclusion

The strong MARK hypothesis can be applied to 89.4% of German two-gender nouns and the weak MARK hypothesis to 92.4% of German two-gender nouns. Among the remaining cases, 0.5% are epicene in the plural, 4.2% are of a lexical nature and 2.2% are words of foreign origin.

We conclude that the MARK is relatively well suited for the German language.

8. The French JOIN Method

The French JOIN (called "écriture inclusive") is a gender-neutral writing method using <·>, <.> or <-> as separator grapheme: <étudiant·e·s>, or <étudiant.e.s>, or <étudiant-e-s> for <étudiants et étudiantes>. It uses the sorting order of gender-specific forms as a criterion for the order of suffixes (cf. §8.3).

8.1. History

Between October 2017 and March 2018, there was an animated debate in the French media concerning gender-neutral writing. The spark that ignited the debate (Manesse and Siouffi, 2019, p. 7) was the publication, on September 22, 2017, in the right-wing daily newspaper *Le Figaro* of the following sentence, taken from a 3rd grade school book (Le Callennec, 2017):

> Grâce aux agriculteurs·rices, aux artisan·e·s et aux commerçant·e·s, la Gaule était un pays riche.[15]

This debate culminated with a statement by the French Academy:

> Prenant acte de la diffusion d'une « écriture inclusive » qui prétend s'imposer comme norme, l'Académie française élève à l'unanimité une solennelle mise en garde. La multiplication des marques orthographiques et syntaxiques qu'elle induit aboutit à une langue désunie, disparate dans son expression, créant une confusion qui confine à l'illisibilité. On voit mal quel est l'objectif poursuivi et comment il pourrait surmonter les obstacles pratiques d'écriture, de lecture – visuelle ou à voix haute – et de prononciation. Cela alourdirait la tâche des pédagogues. Cela compliquerait plus encore celle des lecteurs.
>
> Plus que toute autre institution, l'Académie française est sensible aux évolutions et aux innovations de la langue, puisqu'elle a pour mission de les codifier. En cette occasion, c'est moins en gardienne de la norme qu'en garante de l'avenir qu'elle lance un cri d'alarme: devant cette aberration « inclusive », la langue française se trouve désormais en péril mortel, ce dont notre nation est dès aujourd'hui comptable devant les générations futures.[16] (Académie française, 2017)

The statement about French language being in "mortal danger" seems utterly exaggerated, but may be due to the fact that, until then, the French Academy had been dealing with the acceptability of individual words, and had never to face a meta-technique which applies to tens of thousands of words.

On November 22, 2017, the Prime Minister Édouard Philippe officially prohibited the use of "écriture inclusive" gender-neutral writing in public administration:

> […] je vous invite, en particulier pour les textes destinés à être publiés au *Journal officiel de la République française*, à ne pas faire usage de l'écriture dite

15. "Thanks to farmers$_{gn}$, craftsmen$_{gn}$ and merchants$_{gn}$, Gaul was a wealthy country."

16. "Taking note of the spread of an "inclusive writing" system that claims to become a norm, the French Academy unanimously raises a solemn warning. The multitude of orthographic and syntactic phenomena that it induces leads to a disunited language, disparate in its expression, creating confusion that reaches illegibility. We can hardly identify the goal of this inclusive writing, and we don't see how it could overcome the practical obstacles of writing, reading—visual or aloud—and pronunciation. It would make the task of pedagogues harder. And it would further complicate the task of readers. // More than any other institution, the French Academy is sensitive to developments and innovations in language, since its mission is to codify them. On this occasion, to guarantee the future and to preserve the norm, the French Academy raises an alarm: facing this "inclusive" aberration, the French language is currently in a state of mortal danger, and our nation carries the responsibility of this issue with respect to future generations."

> inclusive, qui désigne les pratiques rédactionnelles et typographiques visant à substituer à l'emploi du masculin, lorsqu'il est utilisé dans un sens générique, une graphie faisant ressortir l'existence d'une forme féminine. Outre le respect du formalisme propre aux actes de nature juridique, les administrations relevant de l'État doivent se conformer aux règles grammaticales et syntaxiques, notamment pour des raisons d'intelligibilité et de clarté de la norme.[17] (Philippe, 2017)

The Prime Minister's reaction to graphemic gender-neutral writing is surprising when we consider the fact that since 2013 his office has been supervising a governmental consulting instance, the *Haut Conseil à l'égalité entre les femmes et les hommes* ("High Council for Equality between Women and Men"), that published in 2016 a guide for gender-neutral writing, including specifications for the French JOIN method.

The appendix of this document (Abily et al., 2016, p. 59–61) contains 96 examples of JOIN gender-neutral forms, using the period <.> as separator grapheme, in singular and plural number. Out of these examples, 15 are epicene. Compared to the classification of French nouns and adjectives that we give in §8.4, the non-epicene examples of (ibid.) are distributed as follows (in %):

Class	1	2	3	4	5	6	7	8	9	10
Ratio in our study	53.3	21.3	6.7	4.6	3	2.6	2.4	2.3	1.7	1.2
Frequency in (ibid.)	28	4	0	5	9	1	16	8	3	3
Ratio in (ibid.)	34.5	4.9	0	6.2	9.4	1.2	19.8	9.9	3.7	3.7

As we see in this table, Class 2 (words with ∅/ne, s/nes pattern, such as <doyen·ne·s>) is underrepresented in Abily et al. (ibid.) and Class 3 (pattern ∅/e/∅/es, as in <acquis·e·s>) is completely absent. The absence of Class 3 may be due to the fact that there is a problem in its representation: in other classes, the semantics of a double-separator expression A·B·C are obtained as follows:

1. to obtain the masculine singular form, read A;
2. to obtain the feminine singular form, read AB;
3. to obtain the masculine plural form, read AC;
4. to obtain the feminine plural form, read ABC,

17. "I invite you, especially for texts intended to be published in the *Official Journal of the French Republic*, not to make use of so-called inclusive writing, i.e., the editorial and typographical practices aiming at substituting for the use of the masculine gender, when used in a generic sense, a spelling revealing the feminine-gender form. In addition to respecting the specific editorial rules of legal texts, state administration must comply with grammatical and syntactic rules, in particular for reasons of intelligibility and clarity of the norm."

but in the case of Class 3, Rule 3 is not satisfied: taking the first and third block, we get *<acquiss> instead of <acquis>.

On the other hand, Classes 5 and 7–10 are overrepresented in (Abily et al., 2016), probably because they have the longest suffixes.

For Class 6, only a single example is given (<nombreux·ses>), which is actually a special case since it exists solely in the plural. The absence of more Class 6 examples may be due to the fact that it is the only case where the feminine form of the word is alphabetically sorted before the masculine one (e.g., <peureuse> < <peureux>), and the authors would rather avoid entering into details about this fact. In any case, the example <nombreux·ses> which is given in (ibid.) is wrong: according to the rules explained on p. 27 of the same document, the correct form should be <nombreuses·eux> (the feminine suffix before the masculine one).

Besides a small inconsistency (<sportif·ve> instead of <sportif·ive>), the (ibid.) appendix contains an important mistake: the plural forms of all examples in Class 5 end with -<al·e·s>, implying that the masculine plural should be -<als>, which is absurd: for example, in the case of <principal·e>, the masculine plural is <principaux> and the feminine plural <principales>, therefore the gender-neutral form should be <principales·aux>, instead of *<principal·e·s>. Our hypothesis that this is a mistake is corroborated by the fact that on p. 27 of the same document, the correct version <territoriales·aux> appears as an example.

An interesting case (and a class per se) is the one of word <tout·e>, having the pattern t/te, s/tes. If we follow the rule that the plural of the gender-neutral form is obtained by adding <·s> to the singular gender-neutral form, then the gender-neutral plural should be <tout·e·s>. But this contradicts the Rule 3 given above: using this form, the plural masculine form would be *<touts> instead of <tous>. Therefore Abily et al. (ibid.) recommend the plural gender-neutral form <tou·te·s>, which is suboptimal because it does not have the same stem as the singular form <tout·e>. In Fig. 2 the reader can see graffiti originating from the French spring 2016 student demonstrations; the author of the graffiti was probably unsure about the right spelling of the plural gender-neutral version of <tout·e>: unable to choose between <TOU·TE·S> and <TOUT·E·S>, e merged the two forms and ended up with a form with three (!!) separator graphemes.

In 2017, a private communication agency published an additional document containing specifications (Haddad and Baric, 2017), this time using the middle dot <·> as separator grapheme. This document provides the following amendments to Abily et al. (2016):

1. the separator grapheme is a middle dot <·> instead of a period <.> (see also §8.2);
2. the error of Abily et al. (ibid.) concerning the plural of words in Class 5 has been partly corrected: the suffixes are correctly written but their order is still wrong (e.g., the erroneous <local·e·s> of Abily

FIGURE 2. Gender-neutral graffiti <TOU·T·E·S CONTRE LA LOI TRAVAIL!> ("ALL$_{gn}$ AGAINST THE *LOI TRAVAIL*!"), picture taken in Grenoble, in September 2018

et al. (ibid.) has become <locaux·ales>, even though the correct form should be <locales·aux>, cf. §8.3);
3. some additional examples from Class 1 are added, raising the total number of examples of nouns and adjectives to 97;
4. a spelling error is introduced: *<administratr·if> instead of <administrat·if>;
5. the JOIN method is extended to entire words instead of merely suffixes, by writing, e.g., <femme·homme> ("man/woman"). In one case a blank space is even included in the second part of the gender-neutral expression: <du·de la>, where the cognitive load is increased since the reader has to realize that the second part of the gender-neutral expression is not simply <de> but also includes the blank space < > and the following word <la>.

In the period 2018–19 several books on gender-neutral writing in French language have been published, including Manesse and Siouffi (2019), a collective linguistic study of the topic, with information about gender-neutral language issues in English, German, Arabic and Korean.

8.2. Choice of the Middle Dot as Separator Grapheme

Haddad and Baric (2017) justify the choice of the middle dot as follows:

> Le point milieu permet d'affirmer sa fonction singulière d'un point de vue sémiotique et par là d'investir « frontalement » l'enjeu discursif et social de l'égalité femmes·hommes.[18] (ibid., p. 9)

What Haddad and Baric (ibid.) probably mean is that they have chosen the middle dot as an unused—and hence totally new in the French-

18. "The middle dot semiotically asserts its specific function and allows a "frontal" investiture of the discursive and social wager of gender equality."

speaking world—typographical sign, so that it can endow the separator grapheme function as its unique *raison d'être*. This is fundamentally different than the grapheme choices in other methods, such as <I> for German MARK, <@>, <*> and <x> for Spanish SINGLE or the slash </> for Greek JOIN, which are all widely used typographical signs, with a multitude of functions.

Indeed, the middle dot (Unicode 0x00B7 MIDDLE DOT) is used mostly in mathematics (for the multiplication operation, binary operations in algebraic structures, etc.), in Catalan (to separate the two <l> when geminated: <l·l>), in Greek (functioning as a semicolon), in Georgian (functioning as a comma) and in Chinese (as a division marker between transliterated foreign words), but has never been used in French[19]. Nevertheless—and this makes it a good choice for a new character to introduce into the French writing system—the middle dot needs no specialized equipment to be inserted into documents: being available in Western-Europe MacRoman and Windows encodings from the beginning, it can be obtained on French MacOS X and Windows keyboard layouts by simple keystroke combinations.

8.3. Order of Suffixes in the French JOIN Method

The order of suffixes in JOIN methods determines whether a given gender-neutral form is ♂-privileging or ♀-privileging. (Abily et al., 2016) chose to apply the following rule: to *lexicographically compare the masculine and the feminine form* and to use that order for suffixes, e.g., <étudiante> lexicographically sorts after <étudiant>, therefore the gender-neutral form is <étudiant·e> (♂♀ order); <territoriaux> comes after <territoriales>, therefore the gender-neutral form is <territoriales·aux> (♀♂ order).

The hitch is that when a noun has dependencies (articles, pronouns, adjectives), according to agreement rules, all dependencies must keep the same suffix order as the noun, e.g., <les agent·e·s territoriaux·ales>, where the noun <agent·e·s> follows the ♂♀ suffix order and therefore the adjective <territoriaux·ales> must follow the same order.

When the noun is epicene, then the adjective is used for suffix-order determination, e.g., in <les fonctionnaires territoriales·aux>, <fonctionnaires> is the noun and therefore the adjective <territoriales·aux> follows its natural ♀♂ suffix order.

Abily et al. (ibid.) do not consider the situation when there are many adjective dependencies of the same epicene noun: is it the closest one that determines suffix order for all the others? the longest one? the one on the left or the one on the right? As a puzzle we can consider the

19. It is used though in some French dialects: Occitan, Franco-Provençal and Gallo.

noun with two adjectives <charmant·e fonctionnaire territorial·e>, what would be its plural number? There are two possibilities:

<charmant·e·s fonctionnaires territoriaux·ales>,
<charmante·ant·s fonctionnaires territoriales·aux>,

depending on whether the suffix order is given by the first or by the second adjective. In the first case, the plural backtrack value is equal to 1 for the first adjective and to 3 for the second, in the second case plural backtrack values are 2 and 4, respectively.

8.4. Evaluation of the French JOIN Method

To investigate the validity of the JOIN hypotheses for the French language, we extracted data on 52,271 non-epicene two-gender nouns and adjectives from the French Wiktionary (version of August 1st, 2019), which we divided into 16 classes. We calculated backtrack separately for the singular and for the plural number (values separated by a comma in the BT column):

Class	Suffixes	#	BT	Typical example
1	∅/e, s/es	27,852	0,1	étudiant·e, étudiant·e·s
2	∅/ne, s/nes	11,132	0,1	doyen·ne, doyen·ne·s
3	∅/e, ∅/es	3,521	0,0	acquis·e, acquis·e·s
4	eur/euse, eurs/euses	2,397	3,4	contrôleur·euse, contrôleur·euse·s
5	al/ale, aux/ales	1,583	0,4	principal·e, principales·aux
6	x/se, x/ses	1,367	4,5	peureuse·eux, peureuses·eux
7	eur/rice, eurs/rices	1,278	3,4	directeur·rice, directeur·rice·s
8	er/ère, ers/ères	1,188	2,3	premier·ère, premier·ère·s
9	if/ive, ifs/ives	895	2,3	attentif·ive, attentif·ive·s
10	∅/le, s/les	621	0,1	actuel·le, actuel·le·s
11	∅/te, s/tes	327	0,1	marmot·te, marmot·te·s
12	eau/elle, eaux/elles	41	3,4	beau·elle, beaux·elles
13	∅/que, s/ques	24	0,1	cyprianenc·que, cyprianenc·que·s
14	c/que, s/ques	19	1,2	opoulenc·que, opoulenc·que·s
15	et/ète, ets/ètes	16	2,3	complet·ète, complet·ète·s
16	∅/se, ∅/ses	10	0,0	bas·se, bas·se·s
	Total	52,271	0.82	

As can be seen in the table, only Classes 6 (in both numbers) and 5 (in the plural), use the ♀♂ order of suffixes: they correspond to merely 5.6% of the total number of words. The value of 0.82 is the weighted average of backtrack values.

8.5. Conclusion

The (strong) JOIN hypothesis has been formulated in such a way that it is valid for all French words, at the expense of potentially high backtrack values. Nevertheless the cases where the backtrack is high are rare when compared to classes such as 1 and 2, for which backtrack is 0. Therefore we observe that the global average backtrack is quite reasonable (less than one grapheme in average) and we conclude that the JOIN is well adapted to French.

8.6. *Rendez-vous pour amant·e·s égaré·e·s*: An Innovative Use of the French JOIN Method

By common consensus, the gender-neutral expression <un·e étudiant·e> ("a_{gn} $student_{gn}$") is generally used with the semantics "a student of either gender". However, once the gender of a given person is known, the consensus is to use the gender-specific version: <Alice est une étudiante and Bob est un étudiant> ("Alice is a $student_{♀}$ and Bob is a $student_{♂}$").

In his 2019 novel *Rendez-vous pour amant·e·s égaré·e·s* ("Appointment for $lost_{gn}$ $lovers_{gn}$") (Abbel, 2019), Éric Abbel uses JOIN gender-neutral writing in an innovative way, namely to *hide* the gender of the two protagonists, called <O> and <U>. The 137-page book contains 248 gender-neutral expressions, mostly articles and pronouns, but also adjectives and nouns, and even a pair of inclusively-combined proper nouns: <Ève·Adam> (p. 31).

It is interesting to note that in 121 cases (almost half of the total number of cases), Abbel disobeys rules of JOIN as stated in Abily et al. (2016): he does not respect the order of <elle·il>, <celles·eux>, <jalouse·loux>, and writes <instituteur·trice> instead of <instituteur·rice>. This shows that even though the *écriture inclusive* method has been adopted by many users of French, the rules of suffix order have not yet reached consensus.

Using the JOIN method, Abbel has avoided gender specification, an otherwise difficult task in French, because of the many agreements between articles, nouns, adjectives and pronouns. This challenge has been raised previously by Garréta (1986), without any graphemic method.

Garreta's achievement is comparable to the notorious Oulipian constraints, cf. Becker (2012, Chap. I.3).

In addition to gender obfuscation, Éric Abbel uses another graphemic method: the name of a persona in the novel is written in the Cyrillic alphabet, and therefore is indecipherable for the average French reader: <Одержимый>, a word meaning "obsessed" in Russian and Ukrainian.

Both strategies (gender-neutral JOIN and Cyrillic alphabet) are purely graphemic since they have no phonetic representation, and Abbel may very well be the first author using them in French literature.

9. Portuguese: JOIN or SINGLE?

Gender-neutral writing in Portuguese seems to be divided between the influence of "sister-language" Spanish, for which the SINGLE method is perfectly well suited, and the use of the JOIN method with the slash separator grapheme, which is recommended by academia and provides better coverage of the Portuguese language.

In her 2006 PhD Thesis on gender identity construction in the Portuguese magazine VIP (Avanço, 2006), the Brazilian linguist Karla Avanço systematically uses the JOIN method with the slash </> as separator grapheme. In an interview she gave in 2013 to a feminist blog,[20] she says:

> Tentei usar uma linguagem inclusiva na minha tese e usei um pouco de tudo, menos o "x" e o "@", porque acho que não cabem nesse tipo de texto. Usei com frequência as duas formas "a/o", separadas por barra ou escrevendo as duas palavras, por exemplo: "as leitoras e os leitores",[21]

where by "x and @" she refers to SINGLE methods using <x> and <@> as replacement graphemes, and by "a/o form" she refers to the JOIN method with the slash as separator grapheme.

In 2009, seven years before the French specification (Abily et al., 2016), the Portuguese governmental commission for citizenship and gender equality *Comissão para a Cidadania e Igualdade de Género* published a guide for gender-neutral language in public administration (Abranches, 2009). This guide contains specifications for a JOIN method for Portuguese using the slash (*barra*) as separator grapheme. It gives 23 examples of gender-neutral forms, 9 of which are epicene, 7 of Class 1 (cf.

20. https://blogueirasfeministas.com/2013/08/16/linguagem-inclusiva-de-genero-em-trabalho-academico/.

21. "I tried to use inclusive language in my thesis and used a little of everything except for the "x" and the "@," because I don't think they are suitable for this type of text. I often used the two "a/o" forms, separated by slash or by writing them entirely, as in "the readers$_{♀}$ and the readers$_{♂}$"."

§9.1), 6 of Class 2 and one of Class 5 (the word <cidadã/o>). Suffix order is not mentioned in the specification, but from the examples given we can infer that the order chosen is the one with the least backtrack:

- <a/o cidadã/o> (♀♂) with backtrack 0 is chosen instead of <o/a cidadão/ã> which would have a backtrack of 2;
- <o/a monitor/a> (♂♀) with backtrack 0 is chosen instead of <a/o monitora/or> which would have a backtrack of 3;
- for examples of Class 1 (as in <a/o médica/o> or <o/a beneficiário/a>), suffix order is random since they have backtrack 0 in both cases.

Despite the existence of this specification, there seems to be no consensus in the Portuguese-speaking world.

In 2012, a short text (Oliveira, Duque, and Weyl, 2012, p. 129–132) contained in a collective work on Women's Law published in Brasilia is devoted to gender-neutral writing and mentions both SINGLE methods (using <@> or <x> as replacement graphemes) and JOIN methods, as in the following sentence:

> Em textos alternativos e informais, e possivel utilizar o "x" ou mesmo um simbolo como o arroba (a+o=@) para destacar que a/o autor/a esta atenta/o para a linguagem que utiliza,[22]

where the authors mention SINGLE methods but in fact use an JOIN method. In the whole text, the SINGLE method is used 5 times, while the JOIN method is used 12 times.

In 2014, a 114-page long "Manual for the non-sexist use of language" (Souza e Silva et al., 2014) published by the government of the Brazilian state Rio Grande do Sul, mentions no graphemic method whatsoever.

Similarly, in 2019, the Brazilian blogger Thaïs Costa arguments[23] against the use of SINGLE methods with <X> and <@> graphemes, but gives no advice about other graphemic methods to use.

9.1. Evaluation

We have investigated the validity of both SINGLE and JOIN hypotheses for the Portuguese language. To that end we have extracted data on 7,799 nouns and adjectives from the Portuguese Wiktionary (as of August 1st, 2019), out of which we have classified 7,700 words into 48 classes, the most important of which are the following eleven:

22. "In alternative and informal texts, it is possible to use an "x" or even a symbol like the at sign (a + o = @) to highlight the fact that the$_{gn}$ author$_{gn}$ is aware$_{gn}$ of the language e uses."

23. https://comunidade.rockcontent.com/linguagem-neutra-de-genero/

Class	Suffixes	#	BT	Typical example
1	o/a, os/as	6,718	1,2	novo/nova, novos/novas
2	∅/a, es/as	472	0,2	observador/observadora, observadores/observadoras
3	ês/esa, eses/esas	107	2,2	cantonês/cantonesa, cantoneses/cantonesas
4	ão/ona, ões/onas	93	2,3	cinquentão/cinquentona, cinquentões/cinquentonas
5	o/∅, os/s	38	1,2	pagão/pagã, pagãos/pagãs
6	ão/ã, ões/ãs	34	2,3	guadrião/guardiã, guardiões/guardiãs
7	∅/∅, s/s	34	0,0	inventariante, inventariantes
8	e/a, es/as	34	1,2	presidente/presidenta, presidentes/presidentas
9	o/∅, es/s	26	1,2	alemão/alemã, alemães/alemãs
10	∅/a, s/as	18	0,1	cru/crua, crus/cruas
11	ão/oa, ões/oas	16	2,3	brolhão/brolhoa, brolhões/brolhoas
	Total	7,590	1.49	

These classes cover 97.3% of the total number of words extracted. Notice that Class 7 is epicene.

Strong SINGLE *Hypothesis*

Classes 1 and 8 are the only ones satisfying the strong SINGLE hypothesis, since, for example, <nov@>, <nov@s> and <president@>, <president@s> are perfectly symmetric. If we add to this the epicene Class 7, we find that 89.7% of the total words in the table satisfy the strong SINGLE hypothesis.

Weak SINGLE *Hypothesis*

Classes 2, 3, 5, 9 and 10 satisfy the weak SINGLE hypothesis: we can write the forms <observador@>, <observador@s> (♀-privileging), <cantones@>, <cantones@s> (♀-privileging), <pagã@>, <pagã@s> (♂-privileging), <alemã@>, <alemã@s> (♂-privileging) and <cru@>, <cru@s> (♀-privileging).

The method does not work for Class 4 (the -ona suffix being too different from the -ão one), for Class 6 (suffixes -ões and -ãs in the plural) and for Class 11 (suffixes -ão and -oa).

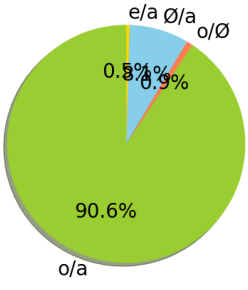
If we add up words satisfying the weak and strong SINGLE hypothesis, we get 98.1% of the words of the table, that is 95.5% of all words extracted from Wiktionary.

We can conclude that the SINGLE method is suitable for Portuguese, even if the large number of irregular forms we found may result in a

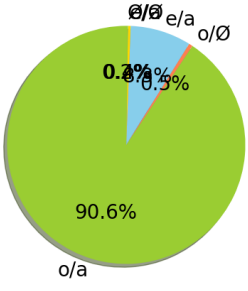
cognitive load for recognizing the grapheme(s) represented by the replacement grapheme.

Polysemy of the REPG

As we can see in the following diagrams, the semantics of the replacement grapheme are quite stable in the singular and in the plural (actually even more than in Spanish): in 90.6% of cases it represents the pair of vowels o/a. The remaining word classes have many suffix pairs (∅/a, o/∅ and e/a for the singular, and e/a, o/∅, o/∅, e/∅ and ∅/a for the plural), the most important being ∅/a for the singular and e/a for the plural. So, like in Spanish, the feminine value of the replacement grapheme is almost always <a>, while the masculine is mostly <o> but can also be empty in the singular or <e> in the plural.



REPG semantics distribution in the singular



REPG semantics distribution in the plural

Strong JOIN *Hypothesis*

The classes represented in the table are all compatible with the JOIN method. Nevertheless the average backtrack we calculated is three times higher than the one for French (§8.4).

9.2. Conclusion

Both methods, SINGLE and JOIN can be used in Portuguese: in the first case, 95.5% of nouns and adjectives of our corpus satisfy the weak SINGLE hypothesis; in the second case, the JOIN hypothesis is satisfied by all words and the average backtrack is higher than the French one, but remains reasonable. The future will show which of the two methods will prevail in Portuguese-language countries.

FIGURE 3. Example of inconsistent use of gender-neutral writing in a page from the official Web site of the University of Athens (`https://www.uoa.gr/foitites/`). The word <φοιτητές> ("students") appears seven times: twice in masculine form and five times in the gender-neutral syllabic separating form <φοιτητές/τριες>. The sentence containing the 7th occurrence has an agreement error: <πολλοὶ/ὲς ξένοι/ες ἀλλὰ καὶ *Ἕλληνες φοιτητὲς/τριες> ("many$_{gn}$ foreign$_{gn}$ but also *Greek$_{♂}$ students$_{gn}$").

10. The Greek JOIN Method

In Greek, gender-neutral language writing ("μὴ σεξιστικὴ γραφή," "non-sexist writing") uses the JOIN method with the slash </> as separator grapheme. The slash—paradoxically called "κάθετος" ("vertical bar") even though it is slanted—has a long history in the Greek language since it is widely used for dates, for law numbers, for administrative codes, as well as for contractions (as in <Δ/νσεις Δ/θμιας Ἐκπ/σης> for <Διευθύνσεις Δευτεροβάθμιας Ἐκπαίδευσης>), therefore its choice as separator grapheme is a natural one.

Compared with JOIN methods in other languages, the Greek JOIN method is more complex to model and evaluate because of declension: in modern Greek there are four cases (nominative, genitive, accusative, vocative) and we therefore have to take into account four cases, two numbers and two genders, resulting in a total of sixteen forms per word.

10.1. History

One of the first Greek publications on gender-neutral language, the 34-page (Tsokalidou, 1996) contains, already in 1996, a total of ten occurrences of forms: nine nouns and one adjective. The classes of these nouns are distributed as follows (with respect to our classification in §10.2): 8 nouns of Class 10, one noun of Class 31, one adjective of Class 3. The order of suffixes seems to be arbitrary.

In 2006 appears the collective work (Pavlidou, 2006) that systematically uses JOIN in all texts. Among them, Makri-Tsilipakou (2006) uses *exclusively* ♀♂ suffix order, while all other texts use ♂♀ suffix order.

In 2016, the Greek Minister of Internal Affairs released an instruction on the "Insertion of the gender dimension in administrative documents" (Kouvela, 2016), mentioning explicitly gender-neutral forms and the JOIN method:

> 1. Συστήνεται ἡ ταυτόχρονη ἀναφορὰ σὲ γυναῖκες καὶ ἄντρες, μέσω τῆς χρήσης καὶ τῶν δύο γραμματικῶν γενῶν, ὅταν τὸ κείμενο ἀναφέρεται σὲ μεικτοὺς πληθυσμοὺς ἢ στὴν περίπτωση ποὺ δὲν προκύπτει τὸ φῦλο.
>
> Αὐτὸ μπορεῖ νὰ ἐπιτευχθεῖ εἴτε μὲ τὴ χρήση ὁλόκληρης τῆς λέξης εἴτε μὲ προσθήκη τῶν καταλήξεων (π.χ. ὁ/ἡ διοικούμενος/η ἢ ὁ διοικούμενος/διοικούμενη, ὁ ἀγρότης/ἀγρότισσα ἢ ὁ/ἡ ἀγρότης/ισσα, οἱ ὑποψήφιοι/ες κ.ο.κ.).
>
> […]
>
> 3. Σὲ περιπτώσεις τῶν οὐσιαστικῶν ποὺ ὁ τύπος τοῦ ἀρσενικοῦ καὶ τοῦ θηλυκοῦ ταυτίζονται, συστήνεται τὰ ἐπίθετα καὶ οἱ ἀντωνυμίες νὰ παρατίθενται καὶ στὰ δύο γένη (π.χ. οἱ διαθέσιμοι/ες ὑπάλληλοι, οἱ ὁποῖοι/ες…).[24]

This ministerial instruction institutionalizes JOIN for the Greek language. It refers to a publication of the General Secretary of Gender Equality, the *Guide of Use of non-Sexist Language in Administrative Documents* (Georgallidou et al., 2018) that contains no specifications, but many

24. "1. When a text is referring to mixed populations or when gender is not explicit, we recommend the simultaneous reference to both men and women by the use of both grammatical genders. This can be achieved either by the use of complete words, or by adding suffixes (e.g., the$_{gn}$ governed$_{gn}$ or the$_{♂}$ governed$_{♂}$/governed$_{♀}$, the$_{♂}$ farmer$_{♂}$/farmer$_{♀}$ or the$_{gn}$ farmer$_{gn}$, the$_{epi}$ candidates$_{gn}$, and son on).//[…]//3. In the case of epicene nouns, adjectives and pronouns should be written in both genders (e.g., the$_{epi}$ available$_{gn}$ employees$_{epi}$ who$_{gn}$…)."

examples of gender-neutral forms. Unfortunately it is full of inconsistencies. For example within a few lines, one can find both forms <ὁμιλήτριες/ές> and <ὁμιλήτριες/τές> (p. 19 lines 13 and −4) differing by the absence of grapheme <τ> in the first case. One also finds forms such as <ὁμιλητριῶν/ῶν>, that make no sense since suffixes are not long enough to be different, breaking an implicit rule of the Join method which is that suffixes should differ (what is actually meant is <ὁμιλητριῶν/τῶν> with a second suffix of length 3).

A subgroup of the authors of (ibid.) also prepared a similar document for the Observatory of Equality of Cyprus. This document, called *Guide for the Transgression of Linguistic Sexism in the Language of Documents of Public Administration of the Republic of Cyprus* (Gkasouka, Georgallidou, and Foulidou, 2016). Similar to (Georgallidou et al., 2018), it contains a multitude of examples, with, again, a lot of inconsistencies.

As for the order of suffixes, Georgallidou et al. (ibid., p. 42) suggest using the ♀♂ order as much as possible:

> Πρόταση: Νὰ χρησιμοποιεῖται συχνὰ ἡ πρόταξη τοῦ θηλυκοῦ γραμματικοῦ γένους. Στόχος εἶναι ἡ ἐπιλογὴ δήλωσης τοῦ γένους/φύλου νὰ εἶναι ἀνατρεπτικὴ ὡς πρὸς τὸν κυρίαρχο γραμματικὸ κανόνα καὶ νὰ λειτουργήσει ἀφυπνιστικά, ὑποδεικνύοντας τὴ δυνατότητα μιᾶς ἐναλλακτικῆς συντακτικῆς διευθέτησης ὄχι ἀμιγῶς γλωσσικῶν ζητημάτων στὴν ἐκπροσώπηση τῶν φύλων στὸ λόγο.[25]

Interestingly, the reason invoked is that "this order is *subversive* with respect to the status quo and can contribute to the *awakening* of the reader," and Gkasouka, Georgallidou, and Foulidou (2016) actually implement this rule throughout the book with the same arguments. Nevertheless, they acknowledge the problem of increased backtrack of the ♀♂ suffix order and add:

> Ὁ Ὁδηγὸς εἶναι μὲν ἐξ ὁλοκλήρου γραμμένος μὲ πρόταξη τοῦ θηλυκοῦ τύπου, μὲ σκοπὸ νὰ δείξει ὅτι αὐτὸ ἀποτελεῖ μιὰ πιθανὴ ἐναλλακτικὴ ἐπιλογὴ τῶν συντακτριῶν/τῶν καὶ γιὰ ἄλλα δημόσια ἔγγραφα, ὡστοσο, τὸ σωστὸ εἶναι πὼς κανένα ἀπὸ τὰ δύο γένη δὲν θὰ ἔπρεπε νὰ δηλώνεται μὲ κατάληξη 3−4 γραμμάτων. Ἐπειδὴ ὅμως αὐτὸ δὲν εἶναι πάντα ἐφικτό, καλὸ εἶναι νὰ ἐναλλάσσονται τὰ γένη ὡς πρὸς τὸ ποιὸ προηγεῖται συντακτικὰ καὶ ποιὸ ἀκολουθεῖ.[26]

25. "Recommendation: To use often the feminine-masculine order in suffixes. The objective we pursue is to have the gender/sex declaration to be subversive with respect to the dominant grammatical rule and to contribute in awakening the reader, illustrating the possibility of an alternative syntactic treatment of not entirely linguistic issues in the representation of gender in discourse."

26. "This Guide is written entirely by using the feminine suffix in the first position, in order to show that this can be a potential alternative author's choice for other public documents. Nevertheless the right way to proceed is by having the suffix of no gender exceed 3−4 letters. As this is not always possible, one should alternate the order of suffixes."

In other words, the authors recommend that writers not follow their practice of systematically privileging the feminine suffix, but rather alternate ♀♂ and ♂♀ suffix orders, not for the sake of gender equality but for practical reasons, as some suffix order may produce very long suffixes.

10.2. Evaluation

In the following we will evaluate the validity of strong and weak JOIN hypotheses for the Greek language and calculate the average backtrack, for both suffix orders.

As it was impossible to extract two-gender nouns from the Greek Wiktionary, we used a different resource: the *Major Greek Dictionary* by Tegopoulos-Fytrakis (Mandala, 1999). From this resource we extracted 15,715 adjectives and 1,033 two-gender nouns. The number of nouns may seem limited, compared for example to those of Wiktionary, but this dictionary does not label words as being both adjectives and nouns, so for example the very common word <φίλος> (“friend”) is labeled only as an adjective.

In the following tables we have classified adjectives and nouns into 37 classes, depending on their decomposition in common prefix and gender-specific suffixes. Here is how to read an entry: in

9	δεξιός	25	ός/ά	οῦ/ᾶς (*ού/άς*)	ό/ά	έ/ά	1.63, 1.38
			οί/ές	ῶν (*ών*)	ούς/ές	οί/ές	

we describe Class 9, a typical example of which is the word <δεξιός>. The number of adjectives in this class is 25. The upper part of the split cells contains singular number suffixes: the suffixes of the nominative case are <ός> for the masculine and <ά> for the feminine word, the suffixes of the genitive case are <οῦ> for the masculine and <ᾶς> for the feminine word, etc. In parenthesized italics, we give the suffixes in the monotonic system, whenever these are different from those of the polytonic system (in this case, they are <ού> and <άς>). The last column contains the average backtracks for ♂♀ and ♀♂ suffix orders (sum of backtracks divided by 16, in this case they are 1.63 for ♂♀ and 1.38 for ♀♂).

10.2.1. Table of Adjectives

Class	Example	#	Nom.	Gen.	Acc.	Voc.	BT
1	φίλος	6,908	ος/η	ου/ης	ο/η	ε/η	1.63, 1.38
			οι/ες	ων	ους/ες	οι/ες	
2	καλός	5,774	ός/ή	οῦ/ῆς (*ού/ής*)	ό/ή	έ/ή	1.63, 1.38
			οί/ές	ῶν (*ών*)	ούς/ές	οί/ές	
3	νέος	1.175	ος/α	ου/ας	ο/α	ε/α	1.63, 1.38
			οι/ες	ων	ους/ες	οι/ες	
4	ψυχοπαθής	891	ής	οῦς (*ούς*)	ή	ής	0, 0
			εῖς (*είς*)	ῶν (*ών*)	εῖς (*είς*)	εῖς (*είς*)	
5	ἀγχογόνος	373	ος	ου	ο	ε	0, 0
			οι	ων	ους	οι	
6	ἀγχώδης	311	ης	ους	η	η	0, 0
			εις	ῶν (*ών*)	εις	εις	
7	γκρινιάρης	196	ης/α	η/ας	η/α	η/α	2.63, 1.63
			ηδες/ες	ηδων/ων	ηδες/ες	ηδες/ες	
8	αὐτουργός	62	ός	οῦ (*ού*)	ό	έ	0, 0
			οί	ῶν (*ών*)	ούς	οί	
9	δεξιός	25	ός/ά	οῦ/ᾶς (*ού/άς*)	ό/ά	έ/ά	1.63, 1.38
			οί/ές	ῶν (*ών*)	ούς/ές	οί/ές	
	Total	15,715				Avg	1.46

10.2.2. Table of Nouns

In this table we use the symbol ↑ whenever the accent of the common prefix of a form is placed one syllable higher than the accent of the common prefix of the other form, e.g., in "ής/↑τρια" the accent of <φοιτητής> is on the ultima of the common prefix, while the accent of <φοιτήτρια> is on the penult of the common prefix. In Class 31 we even have a difference of two syllables between accents, symbolized by the ↑↑ symbol: the masculine <βάτραχος> is accented on the antepenult of the form, which is also the antepenult of the common prefix, while the feminine form <βατραχίνα> is accented on the penult of the form, which is the ultima of the common prefix.

Class	Example	#	Nom.	Gen.	Acc.	Voc.	BT
10	φοιτητής	363	τής/↑τρια	τῆ/↑τριας (*τή/↑τριας*)	τή/↑τρια	τή/↑τρια	2.63, 4.63

			τές/↑τριες	τῶν/τριῶν (*τών/τριών*)	τές/↑τριες	τές/↑τριες	
11	ἐπιβάτης	220	ης/ισσα	η/ισσας	η/ισσα	η/ισσα	1.63, 4.63
			ες/ισσες	ῶν/ισσῶν (*ών/ισσών*)	ες/ισσες	ες/ισσες	
12	ἐπιστάτης	101	της/τρια	τη/τριας	τη/τρια	τη/τρια	2.63, 4.63
			τες/τριες	τῶν/τριῶν (*τών/τριών*)	τες/τριες	τες/τριες	
13	γλωσσάς	44	άς/ού	ᾶ/οῦς (*ά/ούς*)	ά/ού	ά/ού	2.63, 3.63
			άδες/οῦδες (*άδες/ούδες*)	άδων/ούδων	άδες/οῦδες (*άδες/ούδες*)	άδες/οῦδες (*άδες/ούδες*)	
14	δουλευτής	41	ής/↑ρα	ῆ/↑ρας (*ή/↑ρας*)	ή/↑ρα	ή/↑ρα	1.63, 2.63
			ές/↑ρες	ῶν/ρῶν (*ών/ρών*)	ές/↑ρες	ές/↑ρες	
15	καμαριέρης	39	ης/α	η/ας	η/α	η/α	2.63, 1.63
			ηδες/ες	ηδων/ων	ηδες/ες	ηδες/ες	
16	ἰνδιάνος	37	ος/α	ου/ας	ο/α	ε/α	1.63, 1.38
			οι/ες	ων	ους/ες	οι/ες	
17	καφετζής	28	ής/ού	ῆ/οῦς (*ή/ούς*)	ή/ού	ή/ού	2.63, 3.63
			ῆδες/οῦδες (*ήδες/ούδες*)	ήδων/ούδων	ῆδες/οῦδες (*ήδες/ούδες*)	ῆδες/οῦδες (*ήδες/ούδες*)	
18	κλέφτης	27	ης/ρα	η/ρας	η/ρα	η/ρα	1.63, 2.63
			ες/ρες	ῶν/ρῶν (*ών/ρών*)	ες/ρες	ες/ρες	
19	εὐχέτης	15	ης/ις	η/ιδος	η/ιδα	η/ις	1.63, 3.38
			ες/ιδες	ῶν/↑ιδων (*ών/ιδων*)	ες/ιδες	ες/ιδες	
20	ἀρτίστας	13	ας/α	α/ας	α	α	0.38, 0.38
			ες	ῶν (*ών*)	ες	ες	
21	χριστιανός	12	ός/ή	οῦ/ῆς (*ού/ής*)	ό/ή	έ/ή	1.5, 1.38
			οί/ές	ῶν (*ών*)	οί/ές	οί/ές	
22	ξάδελφος	11	ος/↑η	ου/↑ης	ο/↑η	ε/↑η	1.75, 1.63
			οι/↑ες	↑ων/ῶν (*↑ων/ών*)	οι/↑ες	οι/↑ες	
23	τουρίστας	10	ας/τρια	α/τριας	α/τρια	α/τρια	1.63, 4.63

			ες/τριες	ων/τριων (*ών/τριών*)	ες/τριες	ες/τριες	
24	πρίγκιπας	10	↑ας/ισσα	↑α/ισσας	↑α/ισσα	↑α/ισσα	1.63, 4.63
			↑ες/ισσες	↑ων/ισσων (*ων/ισσών*)	↑ες/ισσες	↑ες/ισσες	
25	ἀθεϊστής	9	ής/↑τρια	ῆς/↑τριας (*ής/↑τριας*)	ή/↑τρια	ή/↑τρια	1.63, 4.63
			ές/↑τριες	ων/τριων (*ών/τριών*)	ές/↑τριες	ές/↑τριες	
26	διδάσκαλος	7	↑ος/ισσα	↑ου/ισσας	↑ο/ισσα	↑ε/ισσα	1.75, 4.63
			↑οι/ισσες	↑ων/ισσων (*↑ων/ισσών*)	↑οι/ισσες	↑οι/ισσες	
27	καλλιτέχνης	6	ης/ιδα	η/ιδας	η/ιδα	η/ιδα	1.63, 3.63
			ες/ιδες	ων/↑ιδων (*ών/↑ιδων*)	ες/ιδες	ες/ιδες	
28	συμπέθερος	6	↑ος/α	↑ου/ας	↑ο/α	↑ε/α	1.63, 1.38
			↑οι/ες	ων	↑ους/ες	↑οι/ες	
29	μάγος	6	ος/ισσα	ου/ισσας	ο/ισσα	ε/ισσα	1.88, 4.63
			οι/ισσες	↑ων/ισσων (*↑ων/ισσών*)	ους/ισσες	οι/ισσες	
30	νονός	5	ός/ά	οῦ/ᾶς (*ού/άς*)	ό/ά	έ/ά	1.5, 1.38
			οί/ές	ων (*ών*)	οί/ές	οί/ές	
31	βάτραχος	4	↑↑ος/ινα	↑↑ου/ινας	↑↑ο/ινα	↑↑ε/ινα	1.88, 3.63
			↑↑οι/ινες	↑ων/ινων (*↑ων/ινών*)	↑↑ους/ινες	↑↑οι/ινες	
32	σουρμελής	4	ής/ίδισσα	ῆς/ίδισσας (*ής/ίδισσας*)	ή/ίδισσα	ή/ίδισσα	2.63, 6.63
			ῆδες/ίδισσες (*ήδες/ίδισσες*)	ήδων/ιδισσῶν (*ήδων/ιδισσών*)	ῆδες/ίδισσες (*ήδες/ίδισσες*)	ῆδες/ίδισσες (*ήδες/ίδισσες*)	
33	ἀρραβωνιαστικός	3	ός/ιά	οῦ/ιᾶς (*ού/ιάς*)	ό/ιά	έ/ιά	1.88, 2.63
			οί/ιές	ων/ιῶν (*ών/ιών*)	ούς/ιές	οί/ιές	
34	ἀράπης	3	↑ης/ίνα	↑η/ίνας	↑η/ίνα	↑η/ίνα	2.63, 3.63
			↑ηδες/ίνες	↑ηδων/ίνων	↑ηδες/ίνες	↑ηδες/ίνες	
35	δούκας	3	ας/ισσα	α/ισσας	α/ισσα	α/ισσα	1.63, 4.63
			ες/ισσες	ων/ισσων (*ών/ισσών*)	ες/ισσες	ες/ισσες	
36	αὐτοκράτορας	3	ορας/ειρα	ορα/ειρας	ορα/ειρα	ορα/ειρα	3.63, 4.63
			ορες/ειρες	όρων/↑ειρων	ορες/ειρες	ορες/ειρες	

37	θεράπων	3	ων/αινα	οντος/αινας	οντα/αινα	ων/αινα	4.13, 4.63
			οντες/ αινες	όντων/ αινῶν (*όντων/ αινών*)	οντες/ αινες	οντες/ αινες	
	Total	1,033				Avg	2.19

10.3. Hypotheses

The strong Join hypothesis covers all words, but the difference in accent position between the two forms increases the cognitive load, in particular when the accent of the form of the second suffix is not visible, like in Class 28 for the ♂♀ <συμπέθεροι/ες>, where the reader's brain has to go through all available accent positions for the feminine form: <συμπέθερες> or <συμπεθέρες>? We have the same problem with ♀♂ order in Class 31 <βατραχίνας/ου>: is the masculine form <βάτραχου> or <βατράχου>? The latter issue is due to the phonetic difference between demotic and purified Greek: in purified (and ancient) Greek the antepenult cannot be accented when the ultima is long (genitive <βατράχου>), while this rule is not strict for demotic Greek. Here, the tendency is rather to keep the accent on the same syllable throughout declension (genitive <βάτραχου> for the nominative <βάτραχος>).

For this reason we have introduced a weak version of the Join hypothesis, where accent position in the common prefix is not taken into account.

According to the tables above, there is an accent position difference between forms for Classes 10, 14, 19, 22, 24–29, 31 (difference of two syllables), 34 and 36. These classes represent 47% of nouns but only 2.9% of the total set of nouns and adjectives. We can therefore say that Greek validates the strong Join hypothesis for 97.1% of words, and the weak hypothesis for all words.

As for the calculation of backtrack, we calculated an average of 1.46 for adjectives and 2.19 for nouns, that is a weighted total average of 1.51. This is higher than the backtrack of French, but very close to the value we calculated for Portuguese.

When applying inverse order ♀♂, we obtained a backtrack of 1.23 for adjectives (which is lower than the corresponding value for the ♂♀ order), but a high value of 4.01 for nouns. The total average, due to weighting, gives a backtrack value of 1.41, which is lower than the total average ♂♀ value.

10.4. Conclusion

If we adopt the weak JOIN hypothesis, then all Greek words are covered by this graphemic gender-neutral writing method. The global backtrack is 1.51 in the ♂♀ order of suffixes and 1.41 in the ♀♂ order. These are acceptable values, close to those of French and Portuguese. The difference with other languages is that backtrack can sometimes be very high, but this happens only in rare cases, for which it would be better to write the complete forms in the first place.

11. Experimental Methods

We have mentioned in §7.1 that, in German, the underscore grapheme <_> has been proposed as an alternative to case inversion: <Student_innen> instead of <StudentInnen>. This has been criticized by antibinarist and LGBTQI communities as institutionalizing gender binarity by building a fixed binary (masculine/feminine) marking scheme.

To remedy the rigidity of this underscore usage, which is called *static underscore*, they propose the exact opposite: a *dynamic underscore* ("wandernder Unterstrich" in German: a "wandering underscore") which can be placed anywhere, except at the place where the static underscore would normally be placed. Here is an example (Damm et al., 2014, p. 23):

> We_lche Mita_rbeiterin will denn i_hre nächste Fortbildung zu antidiskriminierender Lehre machen? Sie_r soll sich melden. Der Kurs ist bald voll.[27]

The randomization of the position symbolizes the fact that gender is a continuously changing dynamic process. Besides placing an underscore at random places, the method allows any creative intervention, as in the example above where the feminine pronoun <Sie> has been merged with the masculine pronoun <Er> to give <Sie_r>. In some sense, the idea behind this technique is that it is *not* necessary to produce specific gender-neutral forms, but merely to mark forms in order to show that *gender neutrality is taken into account*. Placing an underscore inside <Mitarbeiterin> shows the gender-neutral *intention* of the writer, and this is enough. If orthography is a set of lexical and morphological constraints, there is no need to add more constraints to the existing ones. On the contrary, gender neutrality provides writers with the opportunity to change the rules, so why not change them in a ludic and creative way?

---

27. "What$_{gn}$ coworker$_{gn}$ would like to have eir training in antidiscriminatory teaching? E should get in touch. The course will soon be full."

From a linguistic point of view, this method is purely graphemic, as it leaves the phonetic realization of words invariant. Similar to the asterisk that denotes ungrammatical forms in linguistics, it denotes gender-neutrality. But contrarily to other signs that *add* information to a word, this one *invalidates* the existing morphological gender-specific information: the reader is invited to ignore the morphological gender mark and to consider the word as being gender-neutral. In other words, it breaks not only the phonetic mechanism (since it has no phonetic realization) but also the morphological one (since the gender morphemes are deliberately ignored).

The wandering underscore has been used in the title <feministische w_orte> of Hornscheidt (2012), a book on gender studies and gender linguistics. According to Damm et al. (2014, p. 24) (which also uses it in its title *W_ortungen statt Tatenlosigkeit!*), the first use of the wandering underscore in a published text was in Tudor (2010), a text on racism and migrationism that was part of a collective book on racism in Germany.

12. Gender-Neutral Forms as Regular Expressions

Regular expressions were introduced in the seminal paper (Kleene, 1951) where Kleene defines finite automata and shows the equivalence between these finite automata and a class of "events" in "nerve nets" (the way neural networks were called at the time), which he calls "regular events"[28]. This was five years before Chomsky, in the equally seminal paper (Chomsky, 1956), defined his hierarchy of formal languages, *regular languages* being the simplest ones, and the only ones that can be described by regular expressions. In 1993, regular expressions became standardized as part of the POSIX family of standards (ISO/IEC, 1993).

Given a set called *alphabet* (in our case: the set of graphemes for a given writing system) and an operator called *concatenation* (in our case: grapheme concatenation), we define *formal words* as concatenations of alphabet members (plus a special word called an *empty word*). A *formal language* is simply a set of formal words. Regular expressions serve to describe (potentially infinite) formal languages by writing paradigmatic words using alphabet members as well as a small number of characters (mostly punctuation) with special semantics. For example, in POSIX notation, `(to|ta){1,3}` represents the formal language of words made out of a single, double or triple concatenation of `to` and/or `ta`, that is the

28. As noted in a footnote in (Kleene, 1951, p. 46), Kleene hesitated to name regular expressions, *prehensible* expressions: "McCulloch and Pitts use a term "prehensible," introduced rather differently; but since we did not understand their definition, we are hesitant to adopt the term."

set of formal words {to, ta, toto, tota, tato, tata, tototo, totota, totato, totata, tatoto, tatota, tatato, tatata}.

As can be seen in the example (to|ta){1,3}, regular expressions provide symbols for separating alternatives (typically the vertical bar | which is equivalent to a Boolean OR), for grouping (typically pairs of parentheses), and for quantifying.

Graphemic gender-neutral writing methods follow, at least partly, the same agenda: to separate alternatives (i.e., gender-specific suffixes) and to group graphemes (i.e., a single-symbol replacement grapheme representing more than one grapheme). If we consider a gender-neutral form as a regular expression, the formal language it represents is exactly the set of gender-specific forms of the word, e.g., in the case of the German StudentInnen, the formal language would be {Studenten, Studentinnen}.

We can therefore legitimately ask the question: can graphemic gender-neutral expressions be represented by regular expressions? If so, this would mean that a simple rule-based decision process would be sufficient to go back and forth from the gender-neutral form to the set of gender-specific forms. To what extent is this possible? In other words, what is the linguistic background necessary to NLP applications to efficiently identify and decode gender-neutral forms?

Let us consider the three types of graphemic gender-neutral writing.

12.1. Single

Translating a gender-neutral form such as the Spanish <nov@s> into a regular expression is straightforward, provided we know the semantics of the <@> grapheme. As we have seen, for 85.9% of Spanish nouns and adjectives, the value of <@> will be o/a, both in the singular and the plural: the corresponding regular expression will be nov(o|a)s. In cases where the masculine suffix is empty, as in <trabahador@>, we will write trabahador(|a). The NLP application will just have to keep a list of word stems for which <@> takes values other than o/a, and use o/a as a default replacement for the rest.

For Italian the process is more complicated since the same <@> will have different values depending on the number of the word. In a sentence like <quest@ bell@ ragazz@ sono pront@>, it is the verb <sono> (in the plural number) that provides the information that the noun and its dependencies are in plural number, and hence <@> takes values i/e in 84.1% of cases. Its translation into a regular expression would be quest(i|e) bell(i|e) ragazz(i|e) sono pront(i|e).

12.2. Mark

Translating a plural Mark gender-neutral form satisfying the Strong Mark hypothesis is straightforward: <StudentInnen> becomes `Student(en|innen)`. As we have seen in §7.2.1, this works for 89.4% of German two-gender nouns.

The situation is more difficult for words satisfying only the Weak Mark hypothesis: in this case, the pair of parentheses of the regular expression has to encompass also the umlauted vowel, as in <JüdInnen> which translates into `J(uden|üdinnen)`[29]. This means that the NLP application will have to detect and process differently umlauted gender-neutral forms[30].

The singular number is more problematic, since case has also to be taken into account: <die StudentIn> (nominative) will be translated `d(er|ie) Student(|in)` while <der StudentIn> (genitive) has to become `d(es|er) Student(en|in)`, since the genitive masculine singular takes its own suffix. The NLP application will have to detect case from the noun's dependencies before translating the gender-neutral expression.

12.3. Join

In the case of Join the main difficulty for translating into regular expressions will be the backtrack.

Indeed, whenever the backtrack is zero, gender-neutral expressions translate straightforwardly: <étudiant·e> becomes `étudiant(|e)`, <étudiant·e·s> becomes `étudiant(|e)s`, etc. In French, this is the case for Classes 1–3, 10, 11, 13 and 16, that is 83.2% of the total number of words represented in Table §8.4. For Portuguese this is the case for Classes 2, 7 and 10, that is only 6.9% of words in Table §9.1. And for Greek this *never* happens, as can be observed in Table §10.2.

For the remaining 16.8% of French words, 93.1% of Portuguese word and 100% of Greek words, the NLP application will have to store the backtrack of each word: when knowing that the backtrack of <traducteur·rice> is 3, it will include 3 graphemes before the separator grapheme into the disjunctive pair of parentheses: `traduct(eur|rice)`.

The situation is more complex for Greek. First of all, the value of backtrack depends on the case of the word: <βάτραχος> (nominative)

29. Writing `J(u|ü)d(en|innen)` seems appealing, but is not correct since the formal language obtained in {Juden, Jüdinnen, Judinnen, Jüden}, where the two last formal words are not German words. This is a superset of the formal language we need, namely {Juden, Jüdinnen}.

30. Except those for which the masculine plural form is also umlauted: the translation of <ÄrztInnen> as `Ärzt(e|innen)` is straightforward.

has a backtrack value of 2, while <βάτραχε> (vocative) has a backtrack value of 1. The NLP application will have to detect the word's case and apply the correct backtrack value.

More difficult is the problem of accent position. In classes containing the ↑ symbol in the table of §10.2, the accent is not in the same position of the stem for both genders. In some cases, both accents are visible: <βάτραχος/ίνα> expands into <βάτραχος/βατραχίνα> and the NLP application will only need to remove the stem accent when adding the already accented suffix. In other cases, the second-gender ("second" in suffix order) accent is not visible: in <βατραχίνα/ος> with a backtrack of 3 the NLP application knows that the first gender-specific form will be <βατραχίνα> and that the second one is made out of the stem <βατραχ> and the suffix <ος>, but the accent is missing. Once again external resources are needed to detect the accent of this word, knowing its gender, number and case.

12.4. A Possible Future of Gender-Neutral Writing

Regular expressions are part of computing, and computing is more and more pervasive in our daily lives. The emergence of graphemic gender-neutral methods may be related to this trend. Indeed, gender-neutral writing is probably one of the first attempts to add regular expression expressive power into natural language (for many years parentheses have been used for that purpose as in <un(e) enseignant(e)>).

The use of regular expressions in natural language has already been explored in poetry, as in

> I need /t(w?o{1,2}) w?r(i|a|ough)te?/.

by American poet Dan Waber, which can be read as "I need to right," "I need to write," "I need two rate," "I need too wrought," and in many other ways (Waber, 2008, p. 149).

Another poem of his may be more difficult to process by the reader:

> /sle[ea]p co-*mes too? (?:me)1,2, but in (?:un|re)fl?its and ?:(?:re)*(?:r?un)?)? (?:st)?art?s\. I sta(?:y|ge) up (?:un)*til the (?:wh?e+)+h?ours\. I c?l(?:a|i)mb (?(?<=amb)or) my s?w(?:ay|eigh) (?:in)?to the bed(?:room)?\. All?one, t?here is (k)?no(?(1)w) goo?d(?:k?night's sle[ea]p)?\./

This kind of poetry will probably not become very popular in the near future, but the idea of using regular expression notation to add expressive power to written natural language may nevertheless lead to new graphemic methods. Simply by adding parentheses to JOIN expressions, one can obtain unambiguous and very creative expressions: <administrat(eur·rice)> contains information about backtrack,

and parentheses can be elsewhere than at word end: <(fe·ho)mmes> (“women or men”), <p(ossi·roba)ble>, etc.

Of course, inclusion of regular expressions or of similar approaches into written language increases cognitive load for writing and reading, and it is debatable whether this approach is efficient for human communication. But as more and more people learn programming languages and regular expressions are omnipresent in them, their emergence in natural language becomes increasingly probable.

13. Conclusion

Gender-neutral writing is a vast subject and we have merely scratched the surface of a particular subarea, namely graphemic methods. After giving a general model of graphemic gender-neutral writing, we have classified the approaches used in French, German, Greek, Italian, Portuguese and Spanish, into three main methods: Single (the simplest method, where a single symbol replaces two different gender-specific suffixes), Mark (where the feminine form is used, suitably marked to show that gender neutrality is meant) and Join (where the masculine and the feminine suffix are both written, separated by a specific grapheme). We have evaluated these methods by their linguistic coverage and the cognitive load required for their use. Finally we have compared graphemic gender-neutral writing methods with regular expressions and have discussed the feasibility of decoding gender-neutral expressions by NLP applications.

References

Abbel, Éric (2019). *Rendez-vous pour amant·e·s égaré·e·s*. Paris: Espaces insécables.

Abily, Gaelle et al. (2016). *Pour une communication publique sans stéréotype de sexe. Guide pratique*. Paris: La documentation française.

Abranches, Graça (2009). *Guia para uma Linguagem Promotora da Igualdade entre Mulheres e Homens na Administração Pública [Guide for an Equality-Promoting Language between Women and Men in Public Administration]*. Lisbon: Comissão para a Cidadania e Igualdade de Género [Governmental Commission for Citizenship and Gender Equality].

Académie française (2017). “Déclaration de l’Académie française sur l’écriture dite «inclusive»”. `http://www.academie-francaise.fr/actualites/declaration-de-lacademie-francaise-sur-lecriture-dite-inclusive`.

Avanço, Karla Fernanda Fonseca Corrêa (2006). "Performatividade e constituição das identidades de gênero na revista VIP [Performativity and Constitution of Gender Identities in the Magazine VIP]". PhD thesis. Universidade Federal de Goiás.

Becker, Daniel Levin (2012). *Many Subtle Channels: In Praise of Potential Literature*. Harvard: Harvard University Press.

Berkins, Lohana (2013). "Nosotres y el lenguaje [We and Language]". https://perma.cc/E2ER-EUCS.

Busch, Christoph (1981). *Was Sie schon immer über Freie Radios wissen wollten, aber nie zu fragen wagten!* Münster: Eigenverlag C. Busch.

Cabral, Mauro, ed. (2009). *Interdicciones, escrituras de la intersexualidad en castellano [Interdictions, writings of intersexuality in Spanish]*. Córdoba: Anarrés Editorial.

Chomsky, Noam (1956). "Three Models for the Description of Language". In: *Transactions on Information Theory* 2, pp. 113–124.

Damm, Anna et al. (2014). *Was tun? Sprachhandeln – aber wie? W_Ortungen statt Tatenlosigkeit!* http://feministisch-sprachhandeln.org/wp-content/uploads/2015/04/sprachleitfaden_zweite_auflage.pdf. AG Feministisch Sprachhandeln der Humboldt-Universität zu Berlin.

Diewald, Gabriele and Anja Steinhauer (2017). *Richtig gendern*. Berlin: Duden.

Gaer, Felice (2009). "Women, International Law and International Institutions: The Case of the United Nations". In: *Women's Studies International Forum* 32.1, pp. 60–66.

García Meseguer, Alvaro (1976). *Sexismo y lenguaje [Sexism and Language]*. Madrid: Cambio 16.

Garréta, Anne (1986). *Sphinx*. Paris: Grasset.

Georgallidou, Marianthi [Γεωργαλλίδου, Μαριάνθη] et al. (2018). *Ὁδηγὸς Χρήσης μὴ Σεξιστικῆς Γλώσσας στὰ Διοικητικὰ Ἔγγραφα [Guide of Use of Non-sexist Language in Administrative Documents]*. Athens: Γενικὴ Γραμματεία Ἰσότητας τῶν Φύλων [General Secretary of Gender Equality].

Gkasouka, Maria [Γκασούκα, Μαρία], Marianthi Georgallidou [Μαριάνθη Γεωργαλλίδου], and Xanthipi Foulidou [Ξανθίππη Φουλίδου] (2016). *Ὁδηγὸς Ὑπέρβασης τοῦ Γλωσσικοῦ Σεξισμοῦ στὴ Γλώσσα τῶν Ἐγγράφων τῆς Δημόσιας Διοίκησης τῆς Κυπριακῆς Δημοκρατίας [Guide for the Transgression of Linguistic Sexism in the Language of Documents of Public Administration of the Republic of Cyprus]*. Nicosia: Παρατηρητήριο Ἰσότητας Κύπρου [Equality Observatory of Cyprus].

Haddad, Raphaël and Carline Baric (2017). *Manuel d'écriture inclusive*. Paris: Mots-Clés.

Haralambous, Yannis [Χαραλάμπους, Γιάννης] (2020). *Ὁδηγὸς γραφῆς γιὰ τὴν ἑλληνικὴ γλώσσα [Writing Guide for the Greek Language]*. Athens: Ἄγρα [Agra].

Hornscheidt, Antje Lann (2012). *Transdisziplinäre Genderstudien*. Vol. 5: *feministische w_orte: ein lern-, denk- und handlungsbuch zu sprache und diskrim-*

inierung, gender studies und feministischer linguistik. Frankfurt a/M: Brandes & Apsel.

ISO/IEC (1993). "ISO/IEC 9945-2:1993 Information Technology – Portable Operating System Interface (POSIX) – Part 2: Shell and Utilities".

Kleene, Stephen C. (1951). *Representation of Events in Nerve Nets and Finite Automata*. Santa Monica, CA: The RAND Corporation.

Kotthoff, Helga and Damaris Nübling (2018). *Genderlinguistik*. Tübingen: Narr Francke Attempto.

Kouvela, Fotini [Κουβέλα, Φωτεινή] (2016). "Ἐγκύκλιος γιὰ τὴν Ἔνταξη τῆς Διάστασης τοῦ Φύλου στὰ Διοικητικὰ Ἔγγραφα [Instruction on the Insertion of the Gender Dimension in Administrative Documents]". Ὑπουργεῖο Ἐσωτερικῶν καὶ Διοικητικῆς Ἀνασυγκρότησης, Γενικὴ Γραμματεία Ἰσότητας τῶν Φύλων, Ἀθήνα, 10/3/2016, Α.Π. 652 [Minister of Internal Affairs, Secretary of Gender Equality, Athens, March 10, 2016, Protocol Number 652].

Le Callennec, Sophie (2017). *Questionner le monde*. Paris: Hatier.

Makri-Tsilipakou, Marianthi [Μακρῆ-Τσιλιπάκου, Μαριάνθη] (2006). "Συμφωνία/Διαφωνία: Ἀλληλεγγύη καὶ Ἀντιπαλότητα στὶς Συνομιλίες Γυναικῶν καὶ Ἀντρῶν [Agreement/Disagreement: Solidarity and Adversity in Discussions between Women and Men]". In: *Γλώσσα - Γένος - Φῦλο [Language - Gender - Sex]*. Ed. by Theodossia-Soula Pavlidou [Θεοδοσία-Σούλα Παυλίδου]. Thessaloniki: Ἰνστιτοῦτο Νεοελληνικῶν Σπουδῶν [Institute of Modern Greek Studies], pp. 81–117.

Mandala, Maria [Μανδαλᾶ, Μαρία], ed. (1999). *Μεῖζον Ἑλληνικὸ Λεξικό [Major Greek Dictionary]*. Athens: Τεγόπουλος Φυτράκης [Tegopoulos Futrakēs].

Manesse, Danièle and Gilles Siouffi (2019). *Le féminin & le masculin dans la langue*. Paris: ESF sciences humaines.

National Council of Culture and Arts [Consejo Nacional de la Cultura y las Artes] (2016). "Guía de lenguaje inclusivo de género [Guide of Gender-Inclusive Language]". https://www.cultura.gob.cl/wp-content/uploads/2017/01/guia-lenguaje-inclusivo-genero.pdf.

Not One Less [Non una di meno] (2017). "Abbiamo un piano. Piano Femminista contro la violenza maschile sulle donne e la violenza di genere [We Have a Plan. Feminist Plan against Male Violence upon Women and Gender Violence]". https://nonunadimeno.files.wordpress.com/2017/11/abbiamo_un_piano.pdf.

Oestreich, Heide (2009). "Die Erektion im Text". taz, March 7, 2009, https://taz.de/Das-Binnen-I-und-die-taz/!5166721/.

Oliveira, Rayane Noronha, Ana Paula Duque, and Luana Medeiros Weyl (2012). "Linguagem Inclusiv@: O que é e para que serve?! [Inclusive Language: What Is It and What Is It For]". In: *O Direito Achado na Rua [Law Found on the Street]*. Vol. 5: *Introdução Crítica ao Direito das Mulheres [Critical Introduction to Women's Law]*. Ed. by José Geraldo de Sousa Ju-

nior, Stefanova Apostolova Bistra, and Lívia Gimenes Dias da Fonseca. Brasília: Universidade de Brasília [University of Brasília].

Patti, Daniela (2018). "El sexismo lingüístico: Un análisis interlingüístico entre español e italiano, con un enfoque particular en el fenómeno del nuevo género neutro en Argentina [Linguistic Sexism: An Interlinguistic Analysis between Spanish and Italian, with a Particular Focus on the Phenomenon of the New Neutral Gender in Argentina]". Tesi di laurea. Università di Bologna.

Pavlidou, Theodossia-Soula [Παυλίδου, Θεοδοσία-Σούλα], ed. (2006). *Γλώσσα - Γένος - Φῦλο [Language - Gender - Sex]*. 2nd ed. Thessaloniki: Ἰνστιτοῦτο Νεοελληνικῶν Σπουδῶν [Institute of Modern Greek Studies].

Philippe, Édouard (2017). "Circulaire du 21 novembre 2017 relative aux règles de féminisation et de rédaction des textes publiés au Journal officiel de la République française". In: *Journal Officiel de la République Française* 272. https://www.legifrance.gouv.fr/eli/circulaire/2017/11/21/PRMX1732742C/jo/texte.

Quilis Merín, Mercedes, Marta Albelda Marco, and Maria Josep Cuenca (2012). *Guía de uso para un lenguaje igualitario (castellano) [Usage Guide for an Egalitarian (Spanish) Language]*. Valencia: Universitat de València.

Robustelli, Cecilia (2012). "Linee guida per l'uso del *genere* nel linguaggio amministrativo [Guidance for the Use of Gender in Administrative Language]". https://www.uniss.it/sites/default/files/documentazione/c._robustelli_linee_guida_uso_del_genere_nel_linguaggio_amministrativo.pdf.

Schoenthal, Gisela (1998). "Von Burschinnen und Azubinnen. Feministische Sprachkritik in den westlichen Bundesländern". In: *Germanistische Linguistik* 139–140, pp. 9–31.

Souza e Silva, Evelise de et al. (2014). *Manual para o uso não sexista da linguagem [Manual on the Nonsexist Use of Language]*. Governo do Estado do Rio Grande do Sul [State Government of South Rio Grande].

Tsokalidou, Petroula [Τσοκαλίδου, Πετρούλα] (1996). *Τὸ φῦλο τῆς γλώσσας [The Gender of Language]*. Athens: Σύνδεσμος Ἑλλήνων Ἐπιστημόνων [Society of Greek Scientists].

Tudor, Alyosxa (2010). "Rassismus und Migratismus: die Relevanz einer kritischen Differenzierung". In: *Rassismus auf gut Deutsch: ein kritisches Nachschlagewerk zu rassistischen Sprachhandlungen*. Ed. by Adibeli Nduka-Agwu and Antje Lann Hornscheidt. Frankfurt a. M.: Brandes & Apsel, pp. 396–420.

Waber, Dan (2008). "Regular Expressions as a System of Poetic Notation". In: *P-Queue* 5, pp. 143–156.