

GENDER IN SOCIETY AND THE GRAMMAR: THE CYCLE OF REPRESENTATIVENESS AND FREQUENCY IN BRAZILIAN PORTUGUESE¹

Raquel Meister Ko. Freitag
Universidade Federal de Sergipe

ABSTRACT. This paper explores the interaction between language and society by proposing an approach through the processing of linguistic variation. Drawing on societal evidence and recent empirical studies in Brazilian Portuguese, it demonstrates how grammatical gender rules are perceived and encoded. The examination of gender agreement rules, exemplified by the choice between *presidente* and *presidenta*, along with results from gender noun frequency studies and the association between gender in society and reference in two-gender common nouns, supports the proposal of a cycle of gender representativeness and frequency in society and its effect on the encoding of grammar rules, reinforcing the mutually constitutive nature of language and society.

Keywords: processing linguistic variation; gender; frequency of usage; representativeness

RESUMEN. Este artículo explora la interacción entre lengua y sociedad, proponiendo un análisis del procesamiento de la variación lingüística. Basándose en evidencias sociales y estudios empíricos recientes en portugués brasileño, demuestra cómo las reglas gramaticales de género son percibidas y codificadas. El examen de las reglas de concordancia de género, ejemplificadas por la elección entre *presidente* y *presidenta*, junto con los resultados de estudios de frecuencia de género de los sustantivos y la asociación entre género en la sociedad y referencia en sustantivos comunes de dos géneros, respalda la propuesta de un ciclo de representatividad y frecuencia de género en la sociedad y su efecto en la codificación de las reglas gramaticales, reforzando la naturaleza mutuamente constitutiva de lengua y sociedad.

Palabras clave: procesamiento de la variación lingüística; género; frecuencia de uso; representación

RESUMO. Este artigo explora a interação entre língua e sociedade, propondo uma análise do processamento da variação linguística. Com base em evidências sociais e estudos empíricos recentes em português brasileiro, demonstra como as regras gramaticais de gênero são percebidas e codificadas. A análise das regras de concordância de gênero, exemplificadas pela escolha entre *presidente* e *presidenta*, junto com os resultados de estudos de frequência de gênero dos substantivos e da associação entre gênero na sociedade e referência em substantivos comuns de dois gêneros, apoia a proposta de um ciclo de representatividade e frequência de gênero na sociedade e o seu efeito na codificação das regras gramaticais, reforçando a natureza mutuamente constitutiva de língua e sociedade.

Palavras-chave: processamento da variação lingüística; gênero; frequência de uso; representatividade

¹ This research was developed within the scope of the projects *Emergence of Non-Binary Gender: Linguistic, Social, and Cognitive Effects* (CNPq Universal 2023 and Internationalization 2023 calls) and *MultiLingualGender* (HORIZON-MSCA-2023-SE-01, European Research Executive Agency).



1. Introduction

The relationship between language and gender is permeated by surface effects that reveal prejudices such as sexism. Are languages sexist? Or are societies sexist and languages “only” codify their sexism? If language and society are mutually constitutive, approaches that explore sexism in language cannot exclude either dimension.

In sociolinguistic and feminist linguistic research, this debate is not new. Lakoff (1973) first argued that linguistic asymmetries mirror social hierarchies rather than cause them, while Cameron (1998) expanded this view by emphasizing the reciprocal influence between linguistic practice and social order. More recent perspectives (Eckert & McConnell-Ginet 2013; Mills 2008) consider linguistic structures and social ideologies to be dialectically intertwined: language both reflects and (re)produces gendered hierarchies. Therefore, addressing sexism in language requires acknowledging this two-way relationship rather than reducing linguistic patterns to mere reflections of social inequality.

This paper returns to the question of sexism in language by presenting another perspective: that of the processing of linguistic variation (Freitag & Soto 2023), an interface field that remains underexplored but has the potential to explain how speakers process language rules based on contexts involving representativeness and frequency.

Since we are dealing with speakers’ perceptions of sexism in language, we rely on indirect evidence of two kinds. One comes from the perspective of language attitudes (Garrett 2010). Attitudes toward linguistic forms are not merely individual preferences; they reflect collective social evaluations. Garrett distinguishes between direct evidence (explicit judgments), indirect evidence (experimental measures), and societal treatment—observable manifestations of attitudes in media, political discourse, educational settings, and institutional norms. Societal treatment is particularly informative in the context of gendered forms, as it reveals how linguistic variants become symbolic sites of ideological negotiation. The other approach comes from descriptive studies that provide data on word frequency and gender associations.

Thus, drawing on societal evidence (Garrett 2010) and results from recent descriptive and experimental studies in Brazilian Portuguese (Pineiro & Freitag 2020, Schwindt 2020, Freitag 2024), we propose a hypothesis about how gender grammar rules are perceived and codified. This perspective posits a cycle of gender representativeness and frequency in society that affects the grammar of the language. We propose a hypothesis about how gender grammar rules are perceived and codified. This perspective posits a cycle of gender representativeness and frequency in society that affects the grammar of the language. We begin by describing the approach to processing linguistic variation (section 2), followed by an analysis of gender rules in Brazilian Portuguese in contexts where one gender dominates others, such as in agreement patterns (section 3).

We then present a case in which two rules vary: the choice between the words *presidente* and *presidenta*, which is driven by social forces of representativeness and frequency that lead to a preference for the two-gender common noun rule over the biform noun rule (section 4). Finally, we mobilize results from empirical studies on noun frequency by gender and on the association between gender in society and occupations denoted by two-gender common nouns to reinforce how the cycle of representativeness and frequency of gender in society plays a role in the codification of gender rules in grammar (section 5).

2. Awareness of linguistic rules

One of the most common definitions of grammar is the set of rules for a language (or variety of a language). These rules are neither universal, univocal, nor stable, as they are sensitive to time, context, and situation. Speakers describe rules from different perspectives, one of which is that of grammatical tradition, which has established a legacy of describing the rules of a prestigious variety in formal or written contexts—what has traditionally been called the standard. This tradition has followed the hegemonic patterns of society, which are reflected in sexism (Carvalho 2021). However, this is not the only possible perspective for describing language rules, especially in situations of plural standards, where different groups of speakers are identified with distinct rules. Where there is more than one standard, there is categorization, hierarchy, and judgment.

Perhaps this is not the most classic definition of the approach to the relationship between language and society, but it emphasizes that the evaluation of language is decisive for the constitution of the linguistic identity of those who use it in terms of social appreciation and awareness. Different approaches have attempted to explain this relationship: Labov (1972), with indicators, markers, and stereotypes; Bell (1984), with audience design; Silverstein (2003), with indexical orders; and Eckert (2008), with indexical fields. These proposals assume that speakers make choices during speech, but these choices are not necessarily at the level of consciousness. That is the level of sociolinguistic awareness, an emergent field in sociolinguistic approaches (Babel 2016; McGowan & Babel 2020; Freitag & Soto 2023).

Sociolinguistic awareness is explicit knowledge that emerges from aggregate experiences of recognizing linguistic differences at the moment of interaction and is understood as linguistically and socially significant. Speakers have an active linguistic awareness – perhaps the best known is phonological awareness, but there are other kinds of awareness, such as grammatical and sociolinguistic awareness.

Sociolinguistic awareness is not the same as conscious knowledge; it means understanding that language is made up of rules without knowing exactly or explicitly what those rules are. The rules are acquired by observing facts in the context of language use through the social frames that each person has built based on their experiences.

Speakers of a language develop linguistic awareness, a range of specific knowledge about it, such as knowledge about its structure and grammar (phonological, morphological, syntactic awareness) or knowledge about social aspects, including linguistic variation, variability in the fit between the speaker and their audience, and the speaker's intentions (sociolinguistic awareness). Knowledge of variation stems from a process during which speakers are exposed to linguistic differences in a situated context, observe them, and systematize the patterns of their use in connection with social facts.

Moreover, linguistic awareness involves knowledge of how language can be deliberately manipulated for persuasive purposes, up to critical language awareness, which involves recognizing how social and discursive patterns are mutually constitutive and how speakers are largely immersed in and conditioned by the linguistic patterns of the community.

The development of knowledge of language is the outcome of observing facts and inferring rules. Even within the same language community, not everyone will infer the same rules by observing the same facts. The value of a rule results from the association between the profiles of the speakers who use it. Rules that are better evaluated with positive value judgments are associated with the profile of speakers with more

education and access to cultural and consumer goods, constituting what is known as a standard. Even if, statistically, this profile of speakers represents a minority in society, their rule becomes hegemonic and supported by prescriptive instruments.

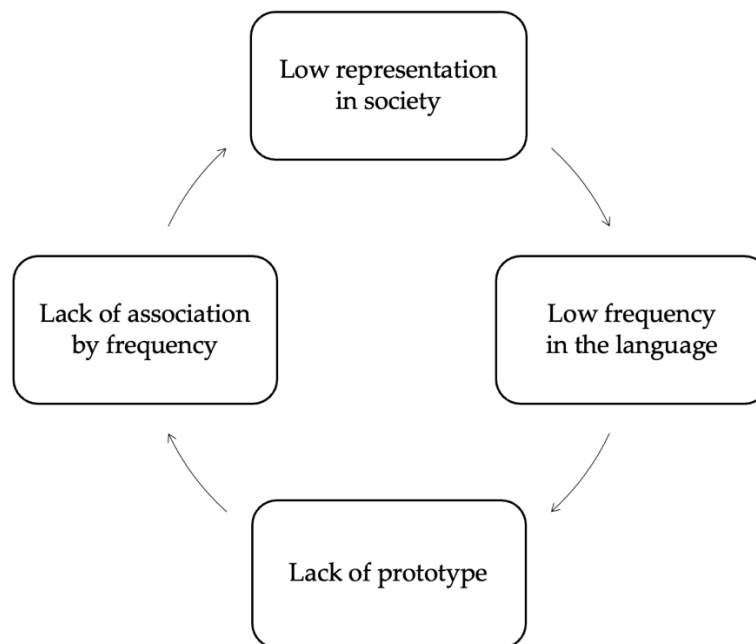
Access to a rule can occur through implicit knowledge, from observation and inference, or through explicit knowledge, via correction (in terms of right and wrong, adequate and inadequate) or even formal instruction. For example, the meaning of plural is implicit knowledge, but the concept of plural as opposed to singular is learned at school, an explicit knowledge.

At some point, speakers mobilize sociolinguistic awareness when trying to explain why someone speaks differently. What do ordinary people and non-linguists think of language? On the one hand, there is popular sociolinguistic awareness: how the speakers perceive, process, and codify the rules of language, including the identification of linguistic varieties as socially or regionally localized markers. On the other hand, there is the force of prescriptivism, a set of normative metalinguistic practices focusing on the value of correctness and "correct" use according to the standard codified in grammar.

While prescriptivism attempts to use scientific models, folk sociolinguistic awareness is still ignored by science, and little is known scientifically about the explanations, claims, and beliefs made by non-experts about the relationship between language and social context. When non-specialists encounter a change in hegemonic rules, their implicit knowledge provides the basis for their explanations, mobilizing the same facts that specialists observe in their descriptions.

By mobilizing sociolinguistic awareness, sociolinguistic processing studies can, for example, provide scientific information for labels attributed to specific social and regional variants by non-linguists as well as help identify the (socio)linguistic features that determine their classification and assist in the task of mapping and classifying varieties and features. The approach to the phenomenon can focus on the description of the rule, how speakers use it, or how speakers understand it.

Figure 1. Cycle of representation and frequency in gender rules



From this perspective of sociolinguistic awareness, we can consider that what is called sexism in language can be explained by a cognitive effect of word usage frequency, which is triggered by the representativeness of its referent in society. Representativeness generates prototypes for associating words with genders, and these prototypes are reinforced through usage frequency in a cycle where low representativeness of people of different genders in spaces of power (or even their absence) leads to low frequency of use of words representing those genders (Fig. 1).

The way in which variable rules are processed depends on the usage frequency of each variant, and the frequency of each variant depends on the social representativeness of the context from which the rule originates.

Through the approach of processing sociolinguistic awareness, we can provide an explanation for sexism in language. Examining gender agreement in Brazilian Portuguese allows us to observe the role of representativeness and frequency in sociolinguistic awareness regarding usage standards in two contexts: when there is variation between two rules (the case of *presidente* and *presidenta*, in section 4), and in the association between nouns and gender (section 5). Before that, section 3 presents the grammatical pattern of gender in Brazilian Portuguese.

3. Gender agreement in Brazilian Portuguese

Gender agreement in Brazilian Portuguese has been presented as a categorical rule, with exceptions in situations of dialectal contact, as in Afro-Brazilian Portuguese, or vacillating gender, as predicted in prescriptive grammar. New forms have emerged in the system to express non-binary gender, such as *-@*, *-x*, and *-e*. However, this was not the focus of this paper (cf. Freitag 2024).

Gender agreement is inflectional and obligatory in Portuguese, a feature reminiscent of Latin cases. It follows that the agreement of both gender and number is redundant (or at least still exists in some varieties of Brazilian Portuguese).

In the grammatical expression of agreement, based on a structuralist approach, Camara Jr. (1975) proposed the artifice of the morpheme \emptyset (zero), which helps demonstrate how gender and number inflections behave according to the concept of marked versus unmarked categories for number and gender.

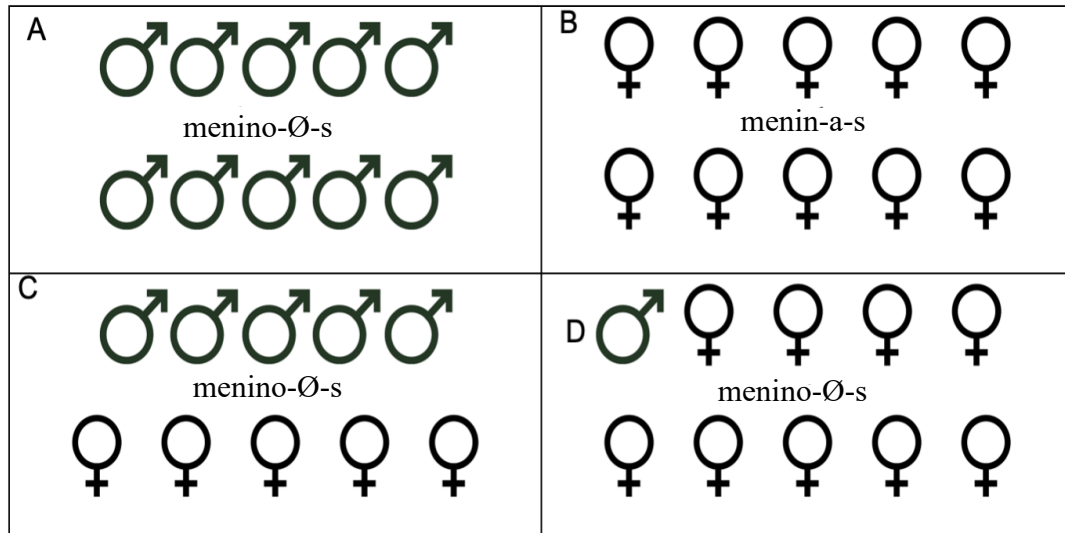
- | | |
|-----|--|
| (1) | O- \emptyset _[gender] - \emptyset _[number] menino- \emptyset _[gender] - \emptyset _[number] bonito- \emptyset _[gender] - \emptyset _[number] |
| | The boy beautiful |
| (2) | O- \emptyset _[gender] -S _[number] menino- \emptyset _[gender] -S _[number] bonito- \emptyset _[gender] -S _[number] |
| | The boys beautiful |
| (3) | A- _[gender] - \emptyset _[number] menin-a _[gender] - \emptyset _[number] bonit-a _[gender] - \emptyset _[number] |
| | The girl beautiful |
| (4) | A- _[gender] -S _[number] menin-a _[gender] -S _[number] bonit-a _[gender] -S _[number] |
| | The girls beautiful |

When referring to people, singular agreement with social gender assumes a binary realization. For example, when saying thank you, a man says *obrigado* (thank you, masculine, $-\emptyset$) and a woman says *obrigada* (thank you, feminine, $-a$). However, when referring to people in a group where there is more than one referent, as well as plural agreement, there is also gender agreement. If the group is homogeneous, as in (2) and (4), gender agreement is not affected in the plural; the plural follows the same gender

pattern as the singular: *os meninos* (the boys) and *as meninas* (the girls). However, in a mixed group where there is more than one boy and more than one girl, the agreement process requires a masculine form.

If the people in a communicative situation are of different genders, one might assume that a quantitative rule could be applied. However, that is not how it works in Brazilian Portuguese; even if the majority are female, the standard for agreement is the masculine gender: this is the rule of the generic masculine.

Figure 2. Patterns of gender agreement in Brazilian Portuguese



In quadrant A (Fig. 2), the set is composed exclusively of male elements, forming a homogeneous group, whereas quadrant B is composed exclusively of female elements, an equally homogeneous group. The agreement marker for referencing A is \emptyset (zero), whereas for referencing B, it is *-a*. In quadrants C and D, there are different proportions of female and male elements: in C, half and half, and in D, only one of the elements is male. For both quadrants, the agreement marker for referencing these sets is \emptyset (zero). The difference between \emptyset in A and \emptyset in C and D is that in A, \emptyset refers to the masculine, while in C and D, \emptyset represents the generic masculine.

Thus, plural masculine gendered nouns have two possible meanings: one as a masculine group and the other as a mixed group. These plurals are ambiguous. For example, in a sentence like *Trabalhadores domésticos deixaram a sala* (The domestic workers left the room), we can interpret *trabalhadores domésticos* as referring to an exclusive group of men or to a group of men and women.

One explanation for these facts in Brazilian Portuguese stems from the reinterpretation of the notion of marked and unmarked, as presented by Camara Jr. (1975). The masculine gender functions as a generic gender or generic masculine. It follows from this reinterpretation that the generic masculine is a form of neuter. This evokes a Latin origin in which there was a class of neuter gender that was incorporated into the masculine in the process of linguistic change leading to Portuguese.

The binarity that produces a marked and defective gender system means that the form expressing masculinity is treated as an unmarked form, the unmarked gender, which is overgeneralized as a generic gender. Other languages follow this same process, including German, which is a gender-marking language but makes use of the generic masculine, which emerges in the same context as the plural of mixed groups.

These processes of the grammaticalization of gender can be explained by internal factors, such as the assimilation of the Latin neuter into the masculine gender and overgeneralization, as historical grammars have traditionally explained. On the other hand, the predominance of one gender over the other configures an arrangement of the linguistic system that conforms to the arrangement of society, and from this perspective represents a form of sexism that permeates the overall process of grammatical tradition development (Carvalho 2021).

To observe how sociolinguistic awareness of gender rules is shaped in specific contexts, the following sections provide societal, descriptive, and experimental evidence showing how representativeness and frequency drive the processes that result in sexism in language.

Therefore, the empirical expression of sexism can be examined in two grammatical contexts: biform nouns and two-gender common nouns (sections 4 and 5), both related to social roles and their relationship to gender representation in society.

4. Biform or uniform nouns: *presidente* or *presidenta*?

Between 2011 and 2016, Brazil was governed by Dilma Rousseff, who chose to inflect her title as *presidenta*, in accordance with federal law and the *Vocabulário Ortográfico da Língua Portuguesa* (VOLP). Federal Law 2,749/1956 regulates the gender of nouns designating public functions.²

Although the word *presidente* is classified as a two-gender common noun, for both masculine and feminine, with the gender specification attributed to the determiner³ (*o* for masculine, *a* for feminine), the *Vocabulário Ortográfico da Língua Portuguesa* (VOLP – Orthographic Vocabulary of the Portuguese Language), which claims to be the official register of words in the Portuguese language and enjoys widespread respect due to its link with the Brazilian Academy of Letters, also registers the form *presidenta*, as well as the use of the suffix *-enta* for feminine positions. Therefore, it is a variable endorsed by normative instruments. Dilma Rousseff chose the *presidenta* form.⁴

However, even before Dilma Rousseff assumed office, public attention focused on linguistic forms rather than solely on political outcomes. From the perspective of societal treatment (Garrett, 2010), the debate surrounding the use of *presidenta* provide empirical evidence of collective language attitudes: it illustrates how speakers evaluate, negotiate, and contest linguistic variants in relation to social ideologies, in this case, gender and authority. Because it is a variable rule, choosing between *presidente* and *presidenta* reflects the way people infer rules from their experiences, their linguistic awareness, and also their stance, including their political position. In the speeches of the 133rd extraordinary deliberative session of the 2nd ordinary legislative session of the 55th legislature, from August 25 to 31, 2016, which addressed the impeachment, the regular expression *a presidenta* appears 432 times, while *a presidente* appears 1,299 times, including in the title of the compilation (Impeachment: o julgamento da Presidente Dilma Rousseff pelo Senado Federal - *Impeachment: the trial of President*

² Available at <https://www2.camara.leg.br/legin/fed/lei/1950-1959/lei-2749-2-abril-1956-355226-publicacaooriginal-1-pl.html>.

³ A determiner is a label applied to a group of words that occupy the leftmost position of a noun phrase. The most frequent and/or most likely class of words to occur in this position in Brazilian Portuguese are articles, but they are not the only ones: demonstrative pronouns, quantifiers, and relative pronouns (possessive pronouns are usually listed as determiners, although there are varieties in which the possessive can be preceded by a determiner and others which cannot).

⁴ Available at: <https://www.correiobraziliense.com.br/app/noticia/politica/2011/03/01/internapolitica.240378/ser-presidente-e-como-escalar-o-everest-todos-os-dias-diz-dilma-na-tv.shtml>

Dilma Rousseff by the Federal Senate). This account suggests that: (1) the rule is effectively variable in its uses, with specific political alignments associated with it; and (2) not even the Brazilian state respected her decision by failing to honor her choice of form.

The Secretariat of Communication of the National Congress (Secom) took a stylistic stance on the subject: while Secom recognizes that both forms are correct, *presidente* is the form they chose.⁵ The media followed the same interpretation of the rule as Secom, referring to the word *presidente*. These examples suggest that when there are two forms, a two-gender common nouns and a specific feminine word, the prescribed rule is to choose the two-gender common nouns. However, this rule has not been consistently followed. In the newspaper *O Globo* on November 12, 2014, two different words were used to designate public functions, with two different rules applied: “*presidente* Dilma Rousseff and *xeica* (sheikha) Moza bint Nasser”. Both *presidente* and *xeique* (sheik) are words endorsed by dictionaries, including the VOLP. But while the feminine word *xeica* is not endorsed, *presidenta* is. Why was the rule applied to Dilma Rousseff that a two-gender common noun (*o/a presidente*), with gender in the determiner, while the rule applied to Moza bint Nasser was a biform noun (*xeique-xeica*)?

This societal evidence shows that different ways of processing variable rules apply to contexts in which, even when a word is a biform noun, its association with a woman in a position of power (still uncommon in society at large) prompts selection of the two-gender common noun rule instead of the biform noun rule. The apparent neglect of the form prescribed by the VOLP and formalized in a 1956 federal law can thus be understood not as a property of grammar but as a reflection of societal evaluations. Given the coexistence of two grammatical options, the prominence of the two-gender common noun (*presidente*) in public discourse suggests that this variant is socially preferred and hegemonic, revealing how the frequency and representativeness of masculine forms in positions of power influence both perception and usage. In other words, societal attitudes toward gendered language – manifested in public debates, media coverage, and everyday commentary – mediate the social codification of grammatical rules, reinforcing the mutually constitutive relationship between language and society.

This way of processing these rules is not unique to Brazilian society. In 2007, Cristina Fernández de Kirchner became president of Argentina, was re-elected in 2011 (she was Dilma Rousseff’s contemporary at a time when South America had two women in the highest positions of power in the region), and ended her term in office on December 9, 2015. Since 2019, she has served as vice president. Before that, Cristina Kirchner had been a member of parliament and a senator. Like Dilma Rousseff, Cristina Kirchner chose the word *presidenta*. Just as in Brazil, her decision was not only disrespected but also scrutinized and delegitimized. When she presided over the first senate session in 2019, one of the senators, José Mayans, addressed her as *presidente* and was corrected by her. The same senator, in the same session, called her *presidente* several times, and she reminded him: “*Presidenta*, Mayans, *presidenta*. Ta, ta.”⁶

Cristina Kirchner repeatedly requested to be addressed as *presidenta*. In 2020, the Royal Spanish Academy (RAE), a far more imposing institution than the Brazilian Academy of Letters (which publishes the VOLP in Brazil) in its role as “guardian of the language”, took a stance on the issue via social media: “#RAEconsultas In reference

⁵ Available at: <https://www12.senado.leg.br/manualdecomunicacao/estilos/presidente-presidenta>.

⁶ Available at: <https://www.lapoliticaonline.com/nota/123688-cristina-reto-a-mayans-porque-le-dijo-presidente-es-presidenta-ta-ta/>

to a woman, the most appropriate option today is to use the form '*presidenta*', a feminine form documented in Spanish since the 15th century and present in the academic dictionary since 1803."⁷

Despite the RAE's position endorsing the rule of gender inflection for designations of public positions, the interpretation favoring the two-gender common noun rule continues to be imposed in Argentine society, regardless of the preference of the person occupying the position.

Notably, Argentina was the first republican country with a presidential system in the world to have a woman as head of state and government. In 1974, after the death of President Juan Domingo Perón (her husband), María Estela Martínez Cartas de Perón assumed the presidency. At the time, the newspaper *La Nación* reported that, in the opinion of the Colombian Academy of Languages, "the correct word" would be *presidenta*.⁸

Cristina Kirchner and María Estela Perón share the fact that they were married to male presidents. Cristina Kirchner's husband, Néstor Kirchner, who died in 2010, was president of Argentina from 2003 to 2007, during which time she served as first lady. María Estela Perón, the third wife of Juan Domingo Perón, was both vice president and first lady. Cristina Kirchner assumed the presidency as a widow and Dilma Rousseff as a divorcée. In none of these cases was there any public discussion about what title would be assigned to the spouses of these women in the highest position of the state.

Almost 50 years ago, the debate over whether the word for a woman in the highest position is *presidente* or *presidenta* was already on the agenda. At the same time that Cristina Kirchner was *residenta* of Argentina, Pilar del Río was *presidenta* of the José Saramago Foundation. In 2008, during an interview, she interrupted the journalist, who had begun to call her *presidente* of the José Saramago Foundation, to clarify that she is *presidenta*, not *presidente*.

Is the difficulty in accepting the form *presidenta* rooted in resistance to the word itself or in reluctance to acknowledge a woman in a position of power? Given that the form has existed for decades in the Latin linguistic world, has circulated for at least fifty years, and is attested in normative instruments such as the VOLP and RAE, attributing the observed "sexism" to linguistic structure alone is untenable.

From the perspective of societal treatment, the variation between *presidente* and *presidenta* reflects the sociolinguistic processing of variable rules: speakers' choices are influenced by their awareness of the rule, the representativeness of the referent in society, and the frequency with which each variant is encountered in context. Public discourse, media coverage, and other societal evaluations act as mechanisms through which these attitudes are expressed and reinforced. Empirical evidence from Brazilian Portuguese further confirms that the frequency of exposure to female-referent forms affects the association of gender with two-gender common nouns, demonstrating how social salience mediates the perception and codification of grammatical rules.

5. Frequency of two-gender common nouns

The following section draws on descriptive studies that provide empirical data on word frequency and gender associations, offering a quantitative foundation for

⁷ Available at: <https://twitter.com/RAEinforma/status/1242541765300826112?s=20>. In the original: #RAEconsultas En referencia a una mujer, la opción más adecuada hoy es usar la forma «presidenta», femenino documentado en español desde el s. XV y presente en el diccionario académico desde 1803.

⁸ Available at: <https://www.pagina12.com.ar/243240-es-presidenta-cristina-la-rae-le-dio-la-razon>. The news item features a facsimile of the newspaper *La Nación* of July 4, 1974, which discusses the word.

understanding how social representation and linguistic variation interact in Brazilian Portuguese.

The frequency of a word in society can provide grammatical information about its social configuration; this is evident in the case of gender in words for occupations expressed by two-gender common nouns in Brazilian Portuguese. While there is an almost categorical association between *-o* and masculine and *-a* and feminine in biform nouns, in nouns that refer to people within the set of two-gender common nouns, the thematic vowel evokes a predominant association with one of the genders for reasons that extend beyond the linguistic system. The regularity of occurrence – the frequency with which the association occurs – configures the prototypicality of gender, and this information is also encoded in grammar. This is evident in the set of two-gender common nouns relating to occupations whose thematic vowel is *-a*, as in *motorista* (driver), *babá* (nanny), *dentista* (dentist), and *frentista* (gas station attendant).

Results from an implicit association task show that in Brazilian Portuguese, the word *motorista* is automatically associated with masculine gender, while *babá* is associated with feminine gender (Pinheiro & Freitag 2020). One explanation for why this occurs is based on experiences with shared community stereotypes associating occupations and genders. In the case of two-gender common nouns, in the absence of grammatical gender information in the nouns themselves, speakers select a gender based on mental representations constructed through socially constructed gender stereotypes.

Stereotypes are representations: a hyper-simplified mental image of a certain category (usually of individuals, institutions, or events) shared in essential respects by a large number of people (Tajfel 1982). In theory, there is no inherently favorable or unfavorable predisposition to a stereotype. The effects of social gender stereotypes can emerge in the gender agreement of nouns in language.

The frequency of two-gender common nouns relating to occupations in Brazilian Portuguese shows marked differences across a substantial number of occupations. In a previous study, a search for 30 occupations revealed more occurrences of masculine gender than feminine gender.

In this study, the frequencies were collected from <https://www.google.com.br> on pages in Brazilian Portuguese using the literal string search option (to differentiate *babá* from *baba* and *camelô* from *camelo*), an option provided by the search engine. The frequency of each occupation was computed by the overall frequency of the item and the frequencies as a female occupation and as a male occupation. Since the query was for strings without morphosyntactic annotation, the search parameters were the occupation word preceded by *uma* (female frequency) and the occupation name preceded by *um* (male frequency). The definite article was not included in the search because in the feminine, *a* could overlap with the preposition *a*; other determiners, quantifiers, or numerals were also not considered, as they do not allow for balanced control. The query was conducted in 2020 (Pinheiro & Freitag 2020).

Babá and *assistente social* (social worker) are the two occupations predominantly associated with feminine gender. *Atendente* (attendant) and *estilista* (stylist) are the two occupations in which the percentage difference between female and male frequencies (range) is lowest, meaning there is no predominance of one gender over the other. In the other 26 occupations, masculine gender predominates.

Table 1. Frequency of genders in occupations in Google search results (Pinheiro & Freitag 2020: 96)

Occupations		N feminine	%	N masculine	%	N total	range
Comandante	Ship captain	3.530	3,5	96.100	96,5	99.630	92,9
Dj	DJ	106.000	3,8	2.720.000	96,2	2.826.000	92,5
Soldado	Soldier	115.000	4,6	2.410.000	95,4	2.525.000	90,9
Motorista	Driver	515.000	5,3	9.210.000	94,7	9.725.000	89,4
Lojista	Shopkeeper	33.900	5,7	564.000	94,3	597.900	88,7
Camelô	Street vendor	3.170	5,8	51.300	94,2	54.470	88,4
Babá	Babysitter	1.500.000	94	95.100	6	1.595.100	88,1
Policial	Policeman	475.000	10,2	4.180.000	89,8	4.655.000	79,6
Fiscal de obra	Building Inspector	11.000	10,5	93.300	89,5	104.300	78,9
Frentista	Shop assistant	19.600	10,6	166.000	89,4	185.600	78,9
Jurista	Lawyer	82.700	11,1	664.000	88,9	746.700	77,8
Eletricista	Electrician	35.200	12,7	241.000	87,3	276.200	74,5
Chofer	Chauffeur	6.190	13	41.300	87	47.490	73,9
Taxista	Taxi driver	98.500	14,5	581.000	85,5	679.500	71,0
Piadista	Joker	2.890	15	16.400	85	19.290	70,0
Agente de trânsito	Traffic agent	57.500	15,4	317.000	84,6	374.500	69,3
Dentista	Dentist	231.000	15,4	1.270.000	84,6	1.501.000	69,2
Servente	Servant	28.200	19,2	119.000	80,8	147.200	61,7
Atleta	Athlete	1.390.000	20,3	5.450.000	79,7	6.840.000	59,4
Assistente Social	Social worker	709.000	79,6	182.000	20,4	891.000	59,1
Gerente de banco	Bank manager	6.020	22	21.400	78	27.420	56,1
Gari	Gari	24.300	23,7	783.00	76,3	102.600	52,6
Pianista	Pianist	403.000	27,7	1.050.000	72,3	1.453.000	44,5
Acrobata	Acrobat	10.600	28,3	26.800	71,7	37.400	43,3
Intérprete	Interpreter	13.900	33,9	27.100	66,1	41.000	32,2
Estudante	Student	3.210.000	34,3	6.140.000	65,7	9.350.000	31,3
Jornalista	Journalist	9.370.000	36,6	16.200.000	63,4	25.570.000	26,7
Repórter	Reporter	2.100.000	42,4	2.850.000	57,6	4.950.000	15,2
Estilista	Stylist	648.000	47	730.000	53	1.378.000	6,0
Atendente	Attendant	1.080.000	51,7	1.010.000	48,3	2.090.000	3,3

The frequency of a two-gender common noun associated with a specific gender is a clue to stereotyping, which can also be observed when comparing the frequencies of nouns and genders in a linguistic sample of speech and in dictionary data, as presented by Schwindt (2020). The samples used were: (i) dictionary léxicon – a survey of all nouns in the *Aurélio Electronic Dictionary*, excluding compounds, derivatives of other nouns, and elements of expressions, totaling 17,049 types; (ii) vernacular use – quantitative analysis of 24 sociolinguistic interviews from the Linguistic Variation in the Southern Region of Brazil Project (VARSUL sample), computing 4,800 tokens. Absolute values were not presented in the tables, nor were independence and association tests conducted; the recalculation was performed using the percentage information provided.

The distribution of biform nouns is proportional, with a slight predominance of feminine nouns over masculine nouns. An independence test conducted only on biform nouns to examine the relationship between gender and data source type showed that the relationship between the two variables is not statistically significant $\chi^2(1, N = 20,540) = 0.0034, p = 0.95$.

Table 2. Nouns by gender in two samples of Brazilian Portuguese (Schwindt 2020: 286 adapted)

	Dictionary sample		Varsul sample		Examples	
	N	%	N	%		
uniform nouns - feminine	8848	51,9	2371	49,	porta	door
uniform nouns - masculine	7348	43,1	1973	41,1	bolo	cake
biform nouns	443	2,6	384	8,0	menino(a)	boy/girl
two-gender common nouns	392	2,3	77	1,6	atleta/estudante	athlete/student

Gendered nouns such as *homem* (man), *mulher* (woman), *menino* (boy), and *menina* (girl) account for 5.5% of the types in the dictionary sample and 13.4% in the VARSUL sample. By gender, masculine nouns predominate over feminine nouns, a statistically significant difference. An independence test conducted only on gendered nouns to examine the relationship between gender and data source type showed that the relationship between the two variables is statistically significant $\chi^2(1, N = 1,246) = 7.9493, p > 0.05$.

These results corroborate the intuitive perception guiding the study: “that there are more masculine words than feminine ones in our lexicon – for some, even, Portuguese is labeled for this, among other reasons, as a macho language” (Schwindt 2020: 279)⁹, as well as illustrating the effects of the salience of the masculine form and its impact on stereotyping. The notion of salience is related to the likelihood of encountering women and men performing functions and tasks related to occupations, and frequency is correlated with the co-occurrence of meanings in participants’ mental lexicons – the more frequent the meaning, the faster it is retrieved (Giora 2003).

Evidence of how traces of beliefs and stereotypes are transposed into the grammar of language is identified in studies involving social gender stereotyping and implicit grammatical gender. Experimental approaches provide clues to the influence of gender stereotypes on language: regarding two-gender common nouns, Garnham, Oakhill, and Reynolds’s (2002) study shows that judgment is more positive for sentences that carry biological characteristics (for example, information about men having beards and only women having babies), while the gender incompatibility effect is negative for sentences that carry characteristics related to clothing. In terms of processing time, sentences related to clothing require more time (suggesting greater processing effort), while sentences with biological characteristics require less time (automatic activation of gender information).

As in English and other languages, in Brazilian Portuguese the effect of gender stereotyping can be observed in two-gender common nouns for occupations, such as *o/a dentista* and *o/a babá*. A study examining judgments of the degree of femininity or masculinity of each of the 30 occupations in Table 1 (“How masculine is this occupation?” or “How feminine is this occupation?”) by Brazilian undergraduates shows that there is an association between the frequency of the occupation on Google and the judgment score of the occupations (Pinheiro & Freitag 2020).

In the most frequent occupations associated with masculine gender, the judgment score is higher, and the same effect is observed with the frequency of feminine gender

⁹ In the original: *que há mais palavras masculinas do que femininas em nosso léxico – para alguns, inclusive, o português se rotula por essa, entre outras razões, como uma língua machista.*

occupations: in the most frequent occupations on Google with feminine gender, the judgment score is also higher.

The results of studies on the correspondence of stereotypes with grammatical gender information suggest that the implicit grammatical gender information in occupations is inferred from the mental representations of gender in stereotypes.

The exploration of two-gender common nouns shows that the inference of grammatical information from the social representation of constructed stereotypes is a salience effect, suggesting that the way speakers experience facts about language is incorporated into grammar at the level of sociolinguistic awareness.

The judgment of the gender of occupations is based on stereotyped roles that are already constructed by collective consciousness in society, such as *eletricista* (electrician) and *soldado* (soldier) for men, and *babá* (nanny) and *estilista* (fashion designer) for women. Exposure to a gender pattern triggers frequency, familiarity, and conventionality, and activates the gender stereotypes associated with the functions and roles performed by each speaker and their occupation in society, as a result of knowledge of the proportions of men and women in these occupations.

On the other hand, lack of representation generates negative stereotyping. The frequency of use of language forms increases as social spaces are occupied; to be included in grammar as a standard, the linguistic form must be recurrent and positively evaluated. The frequency with which the association occurs configures gender prototypicality, and this information is also encoded in grammar. If society is masculine and sexist, language encodes this sexism.

The prototypical frequency of an occupation associated with a gender interferes with the representation in grammar, in the way speakers perceive the rules, in terms of sociolinguistic awareness.

6. Conclusion: usages that reflect sexism?

Drawing on evidence from Brazilian Portuguese, particularly the contextualization of variation in how speakers understand grammatical rules – as in the case of *presidente* and *presidenta*, where the rules of the biform noun (*presidente* and *presidenta*) and the two-gender common noun (*presidente*) coexist – and the association between gender frequency in society and two-gender common nouns for occupations, this paper has sought to demonstrate that examining how speakers perceive gender rules from the perspective of processing linguistic variation helps explain why certain gender reference rules remain uncommon. This reflects the underrepresentation of specific groups in socially prominent positions.

Societal evidence from public debates, media discussions, and institutional discourses surrounding the use of *presidenta* vs. *presidente* reveals collective language attitudes that mirror societal evaluations of gender roles. These social evaluations shape how speakers interpret, adopt, or resist particular linguistic variants, illustrating that linguistic perception is inseparable from social ideologies and practices.

In line with classic feminist linguistic research such as Lakoff (1973) and Cameron (1998), and following approaches that consider linguistic structures and social ideologies to be dialectically intertwined (Eckert & McConnell-Ginet 2013; Mills 2008), the findings discussed here reinforce that language does not originate sexism but participates in its (re)production through the social meanings speakers attach to linguistic choices. Language does not merely codify society's sexism: considering sociolinguistic awareness, linguistic forms become ideological sites where social tensions are manifested. Thus, the use of *presidenta* or the preference for *presidente*

reflects not an inherent property of grammar but rather the reactive stance of social sectors that resist women's presence in positions of power.

The scarcity of frequent exemplars leads to an absence of gender prototypes – or gender generalizations – by association with frequency, creating a cycle of gender representativeness and frequency in society. In this way, the perception of sexism in language results from how gender representativeness in society interferes with the way gender rules are inferred and codified in grammar. Breaking this cycle of women's representativeness and frequency in prominent social positions is, therefore, not merely a sociopolitical act but also a linguistic one, as it denaturalizes the masculine as generic and hegemonic in grammar, reaffirming the mutually constitutive relationship between language and society.

Raquel Meister Ko. Freitag
Laboratório Multiusuário de Informática e Documentação Lingüística (LAMID)
Departamento de Letras Vernáculas (DLEV)
Universidade Federal de Sergipe
Didática II, Av. Marcelo Deda Chagas, s/n - Jardim Rosa Elze, São Cristóvão -
SE, 49107-300, Brasil
rkofreitag@uol.com.br
ORCID: 0000-0002-4972-4320

References

- Babel, A. (2016). *Awareness and control in sociolinguistic research*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139680448>
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204. <https://doi.org/10.1017/S004740450001037X>
- Camara Jr., J. M. (1975). *Estrutura da língua portuguesa*. Vozes.
- Cameron, D. (1998). *The feminist critique of language: A reader* (2nd ed.). Routledge.
- Carvalho, D. S. (2021). *A domesticação da gramática de gênero*. Pontes.
- Eckert, P., & McConnell-Ginet, S. (2013). *Language and gender* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139245883>
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>
- Freitag, R. M. K. (2025). Gender-neutral language in Brazilian federal draft laws. In B. Fagard & A. M. Abrantes (Eds.), *Language and gender* (pp. 66–82). Universidade Católica Editora.
- Freitag, R. M. K., & Soto, M. (2023). Processamento da variação linguística: desafios para integrar aquisição, diversidade e compreensão em um modelo de língua. *Revista de Estudos da Linguagem*, 31(2), 397–431. <https://doi.org/10.17851/2237-2083.31.2.397-431>
- Freitag, R. M. K. (2024). *Não existe linguagem neutra! Gênero na sociedade e na gramática do português brasileiro*. Contexto.
- Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, 30(3), 439–446. <https://doi.org/10.3758/BF03194944>
- Garrett, P. (2010). *Attitudes to language*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511844713>
- Giora, R. (2003). *On our mind: Salience, context, and figurative language*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195136166.001.0001>
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

- Lakoff, R. (1973). Language and woman's place. *Language in Society*, 2(1), 45–80. <https://doi.org/10.1017/S0047404500000051>
- McGowan, K., & Babel, A. (2020). Perceiving isn't believing: Divergence in levels of sociolinguistic awareness. *Language in Society*, 49(2), 231–256. <https://doi.org/10.1017/S0047404519000782>
- Mills, S. (2008). *Language and sexism*. Cambridge University Press.
- Pinheiro, B. F. M., & Freitag, R. M. K. (2020). Estereótipos na concordância de gênero em profissões: efeitos de frequência e saliência. *Revista Linguística*, 16(1), 85–107. <https://doi.org/10.31513/linguistica.2020.v16n1a31637>
- Schwindt, L. C. (2020). Sobre gênero neutro em português brasileiro e os limites do sistema linguístico. *Revista da ABRALIN*, 19(1), 1–23. <https://doi.org/10.25189/rabralin.v19i1.1709>
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3–4), 193–229. [https://doi.org/10.1016/S0271-5309\(03\)00013-2](https://doi.org/10.1016/S0271-5309(03)00013-2)
- Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1–39. <https://doi.org/10.1146/annurev.ps.33.020182.000245>