Gendered Grammar or Ingrained Bias?

Exploring Gender Bias in Icelandic Language Models

Steinunn Rut Friðriksdóttir, Hafsteinn Einarsson
University of Iceland
Sæmundargata 2, Reykjavík
srf2@hi.is, hafsteinne@hi.is

Abstract

Large language models, trained on vast datasets, exhibit increased output quality in proportion to the amount of data that is used to train them. This data-driven learning process has brought forth a pressing issue where these models may not only reflect but also amplify gender bias, racism, religious prejudice, and queerphobia present in their training data that may not always be recent. This study explores gender bias in language models trained on Icelandic, focusing on occupation-related terms. Icelandic is a highly grammatically gendered language that favors the masculine when referring to groups of people with indeterminable genders. Our aim is to explore whether language models merely mirror gender distributions within the corresponding professions or if they exhibit biases tied to their grammatical genders. Results indicate a significant overall predisposition towards the masculine but specific occupation terms consistently lean toward a particular gender, indicating complex interplays of societal and linguistic influences.

Keywords: language models, gender bias, Icelandic

1. Introduction

Over the last few years, there has been a surge in the use and creation of large language models. These models are trained on a myriad of data and scaling laws show that the output quality can be determined by the amount of data that the model is trained on, i.e. more training data should translate to higher quality output (see for instance Hoffmann et al., 2022). However, this increase in data usage can come with a price. Research has consistently demonstrated that as these large language models acquire linguistic proficiency through data-driven learning, they also absorb and replicate the gender bias, racism, religious prejudices, and queerphobia ingrained within the data (Bolukbasi et al., 2016; Abid et al., 2021; Kurita et al., 2019; Venkit et al., 2022).

Following the recommendations of Blodgett et al. (2020), we explicitely define gender bias as the tendency of these models to generate or perpetuate gender stereotypes. This, in turn, can reinforce harmful societal norms, such as by influencing individuals' perceptions regarding the careers or roles accessible to them based on their gender. Another potential harm is the dismissal of individuals that do not conform to these norms, such as when systems that default to binary gender assumptions exclude non-binary and transgender individuals. Similarly, the default use of the generic masculine in gendered languages can lead to feelings of invisibility and marginalization for women and gender minorities. As large language models can be used for applications such as resume filtering, customer sercive or content moderation, biased models can cause direct harm when decisions are influenced by gender assumptions.

Language models trained on Icelandic are no strangers to these challenges. A recent study by Sólmundsdóttir et al. (2022) explored gender bias in two English-to-Icelandic machine translation systems. Gender-neutral English adjectives assumed masculine or feminine forms in Icelandic translations, creating distinct semantic shifts. Adjectives that describe people's characteristics tended to be translated as masculine if the adjectives had a positive connotation (i.a. *smart*, *strong*) but feminine if they had a negative one (i.a. *stupid*, *weak*). The opposite applied to adjectives describing appearance where positive words (such as 'beautiful') were translated as feminine but negative words (such as 'ugly') were translated as masculine. Housework chores were preceded by feminine adjectives ('I am good at cleaning' refers to a woman) but technological and artisan words were preceded by masculine adjectives ('I am good at building houses' refers to a man). This indicates that certain societal ideas about gender are reflected in the models.

It is, however, not always clear where the bias is introduced into models that are trained on highly gendered languages. Existing approaches to gender bias research tend to examine the projection of a word in a gender direction (such as when the projection of the English word 'nurse' leans towards the feminine and is therefore considered to be biased, see for instance Bolukbasi et al. (2016)). These methods tend to be focused on monolingual English models. When languages have a gram-

LREC-COLING 2024, pages 7596–7610
20-25 May, 2024.

matical gender, however, they naturally carry gendered information that is not inherently linked to stereotypes or biases in the real world but rather to morphological agreement.

Gender inflection is fundamental to the Icelandic language as all words that inflect by case (nouns, articles, adjectives, pronouns, and numerals) also have a grammatical gender. The unmarked gender in Icelandic is the masculine form and most occupation terms used for mixed-gendered groups of people in the job market are grammatically masculine. In this paper, we will examine how five Icelandic language models treat occupation words with regard to gender. Specifically, we aim to address the following research questions: Do these models merely echo the gender distribution within respective professions or do they exhibit biases aligned with their grammatical genders? Moreover, do these models potentially manifest unexpected or disproportionate biases toward a particular gender?

We conducted an experiment in which our models were tasked with predicting the appropriate pronoun in a simple sentence, '[he/she/they] is/are a [occupation term].' The outcomes of our study reveal a significant influence of the generic masculine on the training of Icelandic language models, as all of our models exhibit a predisposition towards the masculine when contextual information is absent. For some of our models, this male bias persists even when the sentence is introduced with the phrase 'This is [gendered name],' with specific occupation terms consistently associated with a particular gender, regardless of the assumed gender of the contextual names. Consequently, we infer that the biases embedded in these models defy a singular explanation and instead arise from a complex interplay of societal and linguistic influences as well as training data selection.

2. Background

Societal and linguistic biases are often reflected in language models. DeFranza et al. (2020) researched the appearance of gender bias embeddings trained on Wikipedia and Common Crawl data in 45 languages. Their findings suggest that gendered languages tend to associate men more closely with competence words than women when compared to genderless languages. Despite the fact that Scandinavian countries are often seen as the pinnacle of gender equality, Touileb and Nozza (2022) found that in all of the 9 Norwegian, Swedish, and Danish language models they examined using a translation of the HONEST model toxicity test set (Nozza et al., 2021), sentences about females received more toxic completions than those about males.

Zhou et al. (2019) examined gender bias in Spanish and French word embeddings with regard to both grammatical and semantic gender. They consider both inanimate nouns and animate nouns that usually have two gender forms (like 'doctor' (e. *male doctor*) and 'doctora' (e. *female doctor*). Inanimate nouns were measured using the Word Embeddings Association Test (WEAT, see Caliskan et al., 2017) which measures the association strength of each word with gender concepts. For animate nouns, they tested whether their feminine and masculine forms were symmetrical with respect to gender definition terms. Their findings suggested that inanimate objects were semantically neutral and their gender information simply reflected their grammatical gender. On the other hand, much more gendered information was captured in words referring to occupations, where female occupation words inclined more to the feminine side than male occupation words to the masculine side.

Several studies have been done on the presence and mitigation of gender bias in machine translation models (for an overview, see Savoldi et al., 2021). It is often the case that the data used to train these models contains fewer sentences that refer to women or gender minorities than to men (see for instance Leavy et al., 2020). As we will discuss in Section 7.2, this applies to the training data for at least some of our models as well. Various ways have been used to tackle this problem, such as using gender tags to improve translations (see for instance Vanmassenhove et al., 2019; Elaraby et al., 2018; Stafanovičs et al., 2020) or by using debiasing methods on word embeddings (see for instance Escudé Font and Costa-jussà, 2019). Saunders and Byrne (2020) addressed this problem as one of domain adaption and show significant improvements by using transfer learning on a small set of gender-balanced examples for three language pairs, i.e. by fine-tuning rather than retraining their models from scratch.

So-called bias amplification can happen during training of large language models, resulting in models not only reflecting the biases present within their training data but actually amplifying them (Wang and Russakovsky, 2021; Zhao et al., 2017). Although substantial efforts have been devoted to aligning these language models with societal values (Li et al., 2023), the surge of open-source language models raises pertinent concerns about the adequacy of alignment measures going forward. Even more disconcertingly, the meticulous alignment of models can be subverted through cheap fine-tuning tricks (Yang et al., 2023).

It is, however, important to consider the effect of grammatical gendering on language models beyond the lens of semantic association. Sabbaghi and Caliskan (2022) point out that the magnitude of grammatical gender signals within the models might be different. For instance, they found that the number of grammatically masculine words in training data related to the humanities was consistently higher than in training data related to science. This grammatical gender imbalance could influence their model's tendency to associate humanities with men and sciences with women. In the case of Icelandic, adjectives and other descriptors agree with the object noun and not the subject noun or pronoun in sentences like 'Hún er góður kennari' (e. *She is a good* (masculine form) *teacher*). This grammatical agreement could very well influence gender associations made by our models.

3. Methods

The objective of this study is to investigate the presence of gender bias within language models trained on Icelandic. By focusing specifically on occupation-related terms, we aim to assess whether these models mirror the gender distributions observed in the Icelandic job market. To achieve this, we cross-reference our findings with distribution data obtained from Statistics Iceland. It's important to note that Statistics Iceland provides data in job categories rather than specific occupations. We used 394 occupation terms from the first edition of Ístarf95 (Hagstofa Íslands, 2009), along with a few additions that are also in the second edition, and assigned them to categories to the best of our abilities[1]. The occupation terms along with their assigned categories are listed in the appendix.

3.1. Models

The five models we chose are all variations of the BERT model (Devlin et al., 2019), three of which are trained on Icelandic data exclusively, i.e. `IceBERT` (Snæbjarnarson et al., 2022), `IceBERT-igc` and `IceBERT-ic3`, one is pretrained on Icelandic, but starting from a multilingual checkpoint, i.e. `IceBERT-xlmr-ic3`, and one on Icelandic, Danish, Norwegian, Swedish and Faroese data, i.e. `ScandiBERT` (Snæbjarnarson et al., 2023).

3.2. Evaluating Pronoun Bias in Occupation Predictions

We have the models fill in a masked token referring to a personal pronoun in a sentence including each occupation, i.e. '<mask> is a [occupational term]'[2]. We search through the top 1000 predictions to find the probabilities proposed by the model of the masked token being 'hann' (he) or 'hún' (she)[3]. We then add together the probabilities and normalize them. We refer to the resulting value for the pronoun 'hún' as the feminine pronoun score for a given occupation. For comparison with Statistics Iceland, we group together the occupations determined to be in the same categories and calculate their average gender probabilities.

We additionally calculate the probabilities of each gender for each occupation term when preceded by a gendered name, i.e. 'This is [gendered name]. <mask> is a [occupation]'[4]. We ran this process three times: once with ten traditionally male names, once with ten traditionally female names, and once with names that were determined by us to be gender-neutral, and calculate the average gender score. The purpose is to try and eliminate linguistic bias by providing indicative context as the majority of occupation terms in Icelandic are masculine.

3.3. Statistical Approach

A simple linear regression was conducted to investigate the relationship between two variables by computing the correlation coefficient (r) and the coefficient of determination (R^2). The Wald Test, with t-distribution of the test statistic, was applied to assess the hypothesis that the slope is zero. In addition to the simple linear regression, Spearman's rank correlation was employed to assess the monotonic relationship between the two variables. Spearman's rank correlation coefficient (ρ) is a non-parametric measure that evaluates the strength and direction of the monotonic association between two variables, without assuming linearity or requiring normally distributed data. The significance of the Spearman's rank correlation coefficient was determined using the Spearman's rank correlation test. For all tests, a p-value below 0.05 indicated a rejection of the null hypothesis, suggesting a statistically significant relationship. When computing confidence intervals, we used bootstrapping with 1000 repetitions.

[1] SI's occupational categorization is not publicly accessible and thus we applied our manual categorization of these occupation terms, aligning them with SI's categories as closely as possible. While we cannot guarantee an exact match with SI's categorization, our classification is expected to closely mirror theirs.

[2] The occupational term is provided in the sentence. Only the pronoun should be predicted by the model.

[3] It should be noted that the gender-neutral pronoun 'hán' (singular they) is too infrequent to be represented as a single token by the byte-pair encoding algorithm used by the models and will therefore never appear in their predictions (see Section 7 for further discussion).

[4] Again, the occupational term is provided in the sentence along with the gendered name. Only the pronoun should be predicted by the model.

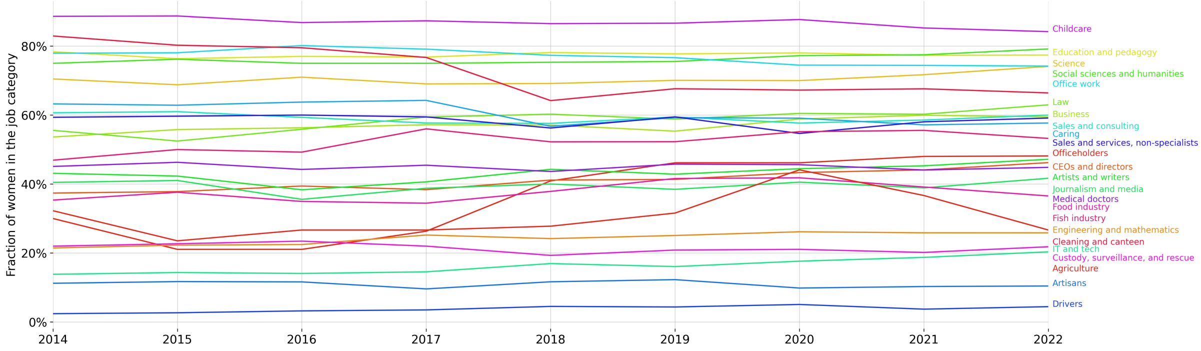

Figure 1: Fraction of women in different job categories over time.

4. Occupational Gender Distribution

Occupational gender distributions from Statistics Iceland provide a relatively clear idea about the role of gender in various fields of work. It's important to acknowledge that these numbers represent a binary perspective and do not encompass information on non-binary individuals. Nonetheless, they effectively illustrate which occupations are or have historically been gendered. Government or city/town council officeholders were twice as likely to be male than female in 2014 but have equalized since then with 48% female employees in the year 2022. A similar trend can be seen with CEOs and management positions although in some fields, such as transportation, still only 28% of management positions are occupied by females. Females hold the overall majority within the medical field but doctors are still males 53% of the time (see Table 1 and Figure 1).

Education and pedagogy are female-centric. In 2014, specialized jobs in the field were female occupied 78% of the time and in 2022, they still held 77%. Similarly, women hold the overwhelming majority of most office jobs, an example being that 75% of employees working as secretaries or consultants were female in both 2014 and 2022. Specialized professionals in the fields of humanities and social sciences have gone from being 75% female in 2014 to an overwhelming 79% in 2022. Women are also more likely to hold positions in accounting, public relations, and record management. Just over 62% of specialized employees in law are women but judges' positions are split evenly.

Conversely, 81% of artisans in 2022 were male, going down from 89% in 2014. Men occupy 80% of engineering and physics research roles. Law enforcement officers are predominantly male, although the proportion of female officers has gone from 11% in 2014 to 27% in 2022. The category of rescue and security workers overall has a 21% rate of female employees. Female motorists (such as truck drivers) increased slightly from 2.4% in 2014 to 4.4% in 2022. Blue collar workers in general are more likely to be male, for instance with 83% of workers in the field of fish processing and transportation being male in 2022, though gender ratios in fish freezing plants are more balanced.

Category	Males	Females
Education	23%	77%
Secretaries	25%	75%
Specialists in law	38%	62%
Government	52%	48%
Doctors	53%	47%
Artisans	81%	19%
Law enforcement	73%	27%
Motorists	96%	4%

Table 1: Example of gender distribution within job categories in 2022 as provided by Statistics Iceland. The table shows selected job categories for illustrative purposes and is not exhaustive.

5. The Generic Masculine and Linguistic Biases

As previously stated, the unmarked gender in Icelandic, like many other gendered languages that derive from the Indo-European language family, is the masculine form. Whereas opponents of language reform argue that this is merely a grammatical convention and that language cannot influence or cause prejudice, only reflect it (see for instance Hjartardóttir, 2004), feminist scholars say that the generic masculine hides women and gender minorities from being seen in historical and contemporary discourse, essentially rendering them invisible (for a review, see Martyna, 1980).

The second wave of feminism (approximately 1960-1980) rejected the idea of women being inherently different from men, stating that the gender roles were rather the result of culture and tradition which needed to be shed. Many women of the movement opposed language that separated the sexes and Icelandic women were no exception. Rather than creating separate occupation terms for women, they wanted women to naturally assimilate any occupation. As a result, women gained access to previously unattainable jobs and gender-neutral, yet masculine occupation terms became popularized for any person working within the corresponding fields. Recently, these views have been brought into question (see for instance Þorbergsdóttir, 2021; Þórhallsdóttir) and there is an ongoing debate about the culturally formative power of the gender-neutral masculine. There has been increasing demand for gender neutralization of Icelandic, particularly with feminists and gender minorities who feel that the use of the masculine form as gender-neutral is exclusionary.

In any case, out of 394 occupation terms used in this experiment, 381 are masculine, 8 feminine[5] and 5 are neuter[6]. In general, when women or non-binary people start working in traditionally masculine fields of work, to this day, they simply adopt the corresponding, grammatically masculine job title. On the other hand, when men start to work in traditionally feminine jobs, Icelandic tends to adopt another, more gender-neutral term for the occupation as a whole or, on rare occasions, the feminine term is kept but an additional, masculine term is adopted alongside it. Examples of the former include when hjúkrunarkona (e. *nurse*, outdated, literally *nursing woman*) was changed to hjúkrunarfræðingur (e. *nurse*, literally *nursing specialist*), a masculine term, and examples of the latter include terms like flugþjónn (e. *male flight attendant*) being adapted alongside flugfreyja (e. *female flight attendant*). This tendency to masculinize occupation terms in order to neutralize them is somewhat paradoxical and may very well be considered to be either a linguistic or a social bias. Most of these terms, however, are semantically gender-neutral in the sense that all genders can have the corresponding titles.

6. Results

6.1. Language and Society

Every one of our models determines jobs in nursing to be more likely to be occupied by females, no matter their specificity or ranking. The modern terms used for these occupations are all grammatically masculine which indicates that the generic masculine does not overshadow the societial fact that nurses are most often women. On the other hand, it does not appear either as if our models are over-predicting female pronouns for these jobs. We do not have the exact gender distribution data from Statistics Iceland for nurses so we work with the assumption that nurses are female in over 90% of cases. Of the five models we studied, only `IceBERT-igc` provided a female pronoun score of 91,93% with the other models providing scores ranging from 68-87%. We think it's safe to assume that for nurses, the models are not over-predicting female pronouns or at least not in any extreme fashion.

Likewise, all of our models determine 'ljósmóðir' (e. *midwife*), 'flugfreyja' (e. *female flight attendant*), 'nunna' (e. *nun*) and 'vinnukona' (e. *female laborer*) to be most likely to be occupied by females in all cases, no matter if they are preceded by masculine or neuter names or not (see Table 4). All of these terms are grammatically feminine and refer either to jobs that are marked, i.e. they are used exclusively for women or have historically referred mostly (or only) to women. Since we do not have the exact job statistics for each individual occupation from SI, only the overarching categories, we cannot make any clear statements regarding over-prediction of female pronouns for these occupations. However, the values are not so extreme that we can easily claim that they are being over-predicted.

The nature of the terms that the models agree to be more likely to be occupied by males is, however, noticeable. Compound words ending with 'maður' (e. *man*) are more likely to be determined to be male occupations, such as 'slökkviliðsmaður' (e. *fire man*), 'veðurathugunarmaður' (e. *weather observer*), 'björgunarsveitarmaður' (e. *rescue worker*) and 'hljóðmaður' (e. *sound engineer*). This, however, does not always apply. The terms 'kjötiðnaðarmaður' (e. *meat worker*), 'kvikmyndatökumaður' (e. *camera man*), 'hárgreiðslumaður' (e. *hair dresser*) and 'héraðsdómslögmaður' (e. *district attorney*) tend to be judged relatively close to being equally likely to be occupied by males or females (although leaning towards male in most cases) and the term 'alþingismaður' (e. *member of pariament*) is judged to have a 64,3% chance of being female by `IceBERT-ic3`.

Compound words ending with 'smiður' (e. *smith*) are also more likely to be labeled as male occupations, such as 'eldsmiður' (e. *blacksmith*), 'skipasmiður' (e. *shipbuilder*), 'rennismiður' (e. *wood turner*) and 'bifreiðasmiður' (e. *auto body builder*). As these all fall under the category of artisans, this is highly representative of actual gender distributions within the field. Similarly, all compound words ending with 'bílstjóri' (e. *driver*) have at

[5]Ljósmóðir (e. *midwife*), eftirherma (e. *impersonator*), nunna (e. *nun*), þerna (e. *maid*), barnfóstra (e. *nanny*), flugfreyja (e. *female flight attendant*), húshjálp (e. *housekeeper*) and vinnukona (e. *female laborer*).

[6]Skáld (e. *poet*), tónskáld (e. *music composer*), leikskáld (e. *play writer*), dagforeldri (e. *daycare provider*), au-pair (e. *au-pair*)

least 60% chance of being judged as male by all of our models, with the notable exception of 'leigubílstjóri' (e. *taxi driver*) having only a 55.4% and 'einkabílstjóri' (e. *personal driver*) a 58.3% chance of being male as judged by ScandiBERT. This aligns fairly well with the gender distributions provided by Statistic Iceland.

Compound words ending with 'meistari' (e. *master*), such as the words 'rafvirkjameistari' (e. *master electrical mechanic*), 'trésmíðameistari' (e. *master carpenter*) and 'matreiðslumeistari' (e. *master chef*) are deemed to be more likely to be male by all of our models. This has two notable exceptions where some of our models will judge 'hárgreiðslumeistari' (e. *master hairdresser*) and 'kjólameistari' (e. *master dressmaker*) as more likely to be female. This is in line with the artisanal nature of these positions and the perceived femininity of the hairdressing and dress-making positions which are most likely societal in nature. In contrast, compounds ending with 'fræðingur' (e. *specialist*, most often used in terms related to academic occupations) do not show a notable trend towards a specific gender outside of the general trend of all occupations being more likely to be judged as male. The most notable exception to this is any term ending with 'hjúkrunarfræðingur' (e. *nurse*, literally *nursing specialist*) which as previously noted is always judged as more likely to be female.

Managerial posts and CEO positions will often be represented by compound words ending with 'stjóri' (e. *boss/chief*). Notably, every single one of these positions is judged as more likely to be male by IceBERT-xlmr-ic3. The other models have a less pronounced bias. ScandiBERT judges 25 out of these 43 positions to be more likely to be female, including terms such as 'leikhússtjóri' (e. *theater director*), 'ráðuneytisstjóri' (e. *permanent secritary of state*, 'gæðastjóri' (e. *quality manager*) and 'starfsmannastjóri' (e. *staff manager*). Managerial positions related to the field of nursing are usually judged as more likely to be female and principals of kindergartens are usually judged to be female. Notably, out of the 25 female managerial positions, as determined by ScandiBERT, two are prefixed with 'aðstoðar' (e. *assistant*), and that prefix is also visible in the results of IceBERT and IceBERT-igc. On the other hand, IceBERT-ic3 seems to place a feminine value on managerial positions related to taxes and customs duty. We do not currently have an explanation for this.

Occupation terms in the field of education show an interesting trend in our models. Most of the models will have a hierarchy where lower-ranking university positions (such as 'aðjúnkt' (e. *adjunct*), lektor (e. *lecturer* and 'dósent' (e. *assosiate professor*), note that 'lektor' is judged to have a 91.9% chance of being male by ScandiBERT) will have a higher chance of being female but the highest-ranking position of 'prófessor' (e. *full professor*) has at least 80% chance of being occupied by males as judged by all models. Other school levels have a less clear but yet similar trend where 'leikskólakennari' (e. *kindergarten teacher*), 'grunnskólakennari' (e. *grade school teacher*) and 'framhaldsskólakennari' (e. *secondary school teacher*) are usually at least equally likely to be female but 'háskólakennari' (e. *university teacher*) is more likely to be male. 'Ökukennari' (e. *driving instructor*) and 'flugkennari' (e. *flight instructor*) are also more likely to be judged as male.

6.2. Overall Trends

In general, as seen in Figure 2, all of our models have a masculine bias when not preceded by context involving a gendered name. As an overwhelming proportion of the occupation terms are grammatically masculine, we suspect that this bias is due mainly to the language itself as only the grammatically feminine terms will have a higher likelihood of being female. Adding a context involving a gendered name would, in theory, eradicate that bias completely, where all female names would result in every job being judged as female, all male names would result in every job being judged as male, and adding all gender-neutral names would have a similar effect as having no context at all[7]. This, however, is not entirely replicated in our results. No model will have 100% chance of all occupation terms being either male or female despite the preceding context. The terms causing the most conflict in these cases are the ones that are heavily biased, either due to linguistic or social properties as has been indicated in this section.

However, our models do not exhibit uniform responses when gendered names are used as context. ScandiBERT demonstrates the most pronounced effect, where the presence of all male names in the context significantly increases the likelihood of generating a corresponding masculine pronoun, approaching nearly 100%. IceBERT-xlmr-ic3 and IceBERT-igc exhibit a similar but less prominent pattern, indicating a reasonably strong comprehension of gender. In contrast, IceBERT-ic3 and IceBERT display a far less noticeable effect. We observe that most of the

[7] It is important to note that we may not actually want the models to have a 100% chance of predicting a feminine pronoun for a perceived female name. With the establishment of the laws on gender autonomy no. 80/2019, a person has the legal right to change their gender registration to fit their gender identity. This implies that names are no longer tied to any specific gender.

models have a masculine bias where the feminine pronoun score is around 0.31. However, IceBERT-xlmr-ic3 is an outlier in this regard, as its mean feminine pronoun score is around 0.22 and its 95% confidence interval does not overlap with the other models.

6.3. Statistical Analyses

The relationship between the language model feminine pronoun score and the mean fraction of women in specific job categories was investigated using both linear regression and Spearman's rank correlation. As the language models are trained on data from different times we decided to average the fraction of women over the time period reported by Statistics Iceland (2014-2022). The results varied depending on the language model used with the overall results shown in Table 2. For the `IceBERT-xlmr-ic3` model we observed the strongest relationship between the language model score and actual gender job proportions according to the correlation coefficient, see Figure 3. For `IceBERT`, the correlation coefficient (r) was 0.40 ($p = 0.056$), and Spearman's rank correlation coefficient (ρ) was 0.36 ($p = 0.086$), both indicating a non-significant relationship.

In contrast, `IceBERT-ic3` demonstrated significant results for both r (0.46, $p = 0.025$) and Spearman's ρ (0.42, $p = 0.041$). `IceBERT-xlmr-ic3` also yielded significant results, with an r of 0.52 ($p = 0.010$) and a Spearman's ρ of 0.45 ($p = 0.027$).

`IceBERT-igc` showed a significant r of 0.50 ($p = 0.012$) but a non-significant Spearman's ρ of 0.33 ($p = 0.110$). Finally, `ScandiBERT` exhibited significant results for both r (0.49, $p = 0.016$) and Spearman's ρ (0.58, $p = 0.003$), with the latter indicating a strong monotonic relationship.

These findings suggest that the choice of language model influences the observed relationship between the feminine pronoun score and the mean fraction of women. While some models showed significant linear relationships (e.g., `IceBERT-ic3`, `IceBERT-xlmr-ic3`, `IceBERT-igc`, and `ScandiBERT`), others did not (i.e., `IceBERT`). Additionally, the strength of the monotonic relationship, as measured by Spearman's rank correlation, varied across models, with ScandiBERT demonstrating the strongest association.

To compare models directly, we computed correlation coefficients that can be interpreted as the pairwise agreement of how language models assign feminine pronoun scores to occupations, see Figure 4 (for Spearman's rank correlation coefficients see Figure 6 in the Appendix). `IceBERT`, `IceBERT-igc` and `IceBERT-xlmr-ic3` show the highest agreement amongst themselves whereas `ScandiBERT` has a lower agreement, which might be explained by the fact that it was trained on Scandinavian languages. Surprisingly, we observe that `IceBERT-ic3` has the lowest agreement with the other models. Perhaps, this is due to the fact that it seems to be worse at determining pronouns from context than the other models as is evident from Figure 2.

Figures 3 and 5 show that most of the occupation terms will be determined to be more likely to be male. The outliers are occupation categories that have a highly gendered connotation, such as jobs in childcare.

Model	r	ρ
`IceBERT`	0.40 (0.056)	0.36 (0.086)
`IceBERT-ic3`	0.46 (0.025)	0.42 (0.041)
`IceB.-xlmr-ic3`	0.52 (0.010)	0.45 (0.027)
`IceBERT-igc`	0.50 (0.012)	0.33 (0.110)
`ScandiBERT`	0.49 (0.016)	0.58 (0.003)

Table 2: The correlation coefficients (r), Spearman's rank correlation coefficients (ρ) and significance levels in brackets for different models when evaluating their relationship to job market data from Statistics Iceland.

7. Discussion

7.1. Gendered Language

It is worth considering what effect gendered language can have on language models such as the ones that have been discussed in this paper. As previously discussed, all of the models show a masculine bias when occupational terms are not preceded with context. Perhaps evident of the proposed erasure of non-male genders, the gender-neutral pronoun 'hán' is infrequent enough to not be represented as a single token by the byte-pair encoding algorithm. Incorporating gender-neutral pronouns would thus require us to predict two mask tokens instead of one, but could be done through an extension of the methodology. It should be noted that this pronoun, along with a few others used for the same purpose, has only existed in the language since approximately 2010 and a large part of the training data for these models is older than that. Additionally, this pronoun refers to non-binary people specifically and does therefore not refer to people of unknown gender in the same way as the singular they in English and as non-binary individuals are a minority group, their mere scarcity may affect the models' treatment of the word.

7.2. Biased data

Out of the five models we evaluated in this paper, three of them (`IceBERT`, `IceBERT-igc` and `ScandiBERT`) were trained at least partially on the

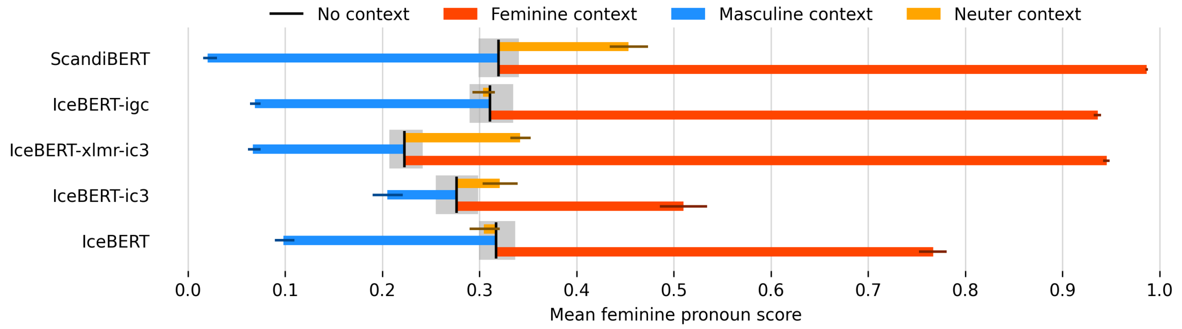

Figure 2: Relative score of a feminine pronoun across language models averaged over all occupations where the score refers to occurrences of a feminine pronoun in a sentence of the form '___ is a X' compared to other pronouns where X refers to an occupation. The vertical black line represents the score without using context of the form 'This is X' where X refers to a feminine, masculine, or a neuter name. The bars represent deviations from the score without context when context is used. The horizontal lines and the gray rectangles are 95% confidence intervals.

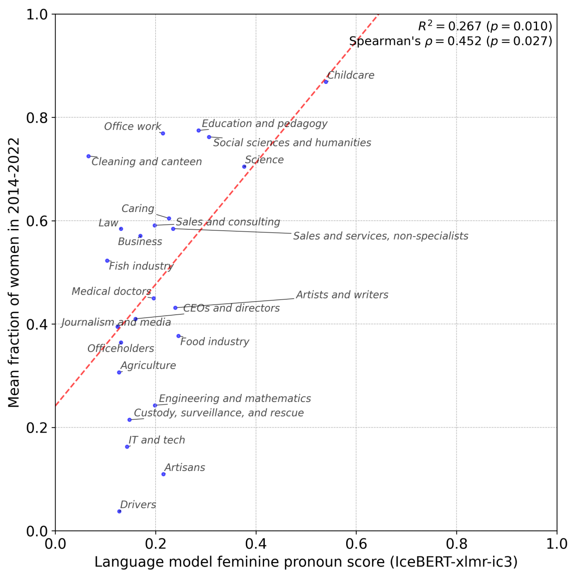

Figure 3: Relationship between feminine pronoun scores (without context) and the fraction of women in different job categories averaged over the period 2014-2022.

	Masc	Fem	Neut
Proper names	52,4%	30%	17,6%
Pronouns singular	29%	14,5%	56,5%
Pronouns plural	55,3%	16,6%	28,1%

Table 3: Grammatical gender distribution in the Icelandic Gigaword Corpus, used to train three of our models.

Icelandic Gigaword Corpus (Steingrímsson et al., 2018). This is one of the largest corpora available for Icelandic and it has been lemmatized as well as tagged with morphosyntactic information. We took advantage of the tagging to examine the gender distributions of proper nouns and personal pronouns in the corpus. As evident by Table 3, the masculine form is a lot more common than the feminine and the neuter except in the case of singular pronouns which is explained by *það* (e. *it*) being counted as well as the gender neutral *hán* (e. singular *they*).

In Figure 2, we see that without context, these three models somewhat reflect the gender distribution of the training corpus with a relative score of approximately 30% for feminine pronouns predicted. We stress that while we cannot conclude that this is the sole reason for our findings, it raises the important issue of data selection bias. It is fundamental that the data used to train models is diverse and includes vocabulary representative of minority groups as overrepresentation of certain groups can lead to others not being represented at all.

7.3. Prestige and Gender Equality

In his paper, Oddsson (2016) reports that while older research has indicated that Icelanders perceive their society as relatively classless, class division increased significantly during neoliberal ascendancy from 1995 to late 2008, where globalization increased the number of low-wage immigrant workers and concentrated income and wealth gave rise to a new super-rich class. Oddsson argues that many Icelanders experienced the 2008 financial collapse as corrective, leaving the derailed 'New Iceland' behind and returning to an egalitarian 'Old Iceland'.

While Icelanders do not seem to consider themselves as a class-divided society, demands for greater social justice are still common. Bernburg and Ólafsdóttir (2023) asked over 1200 partici-

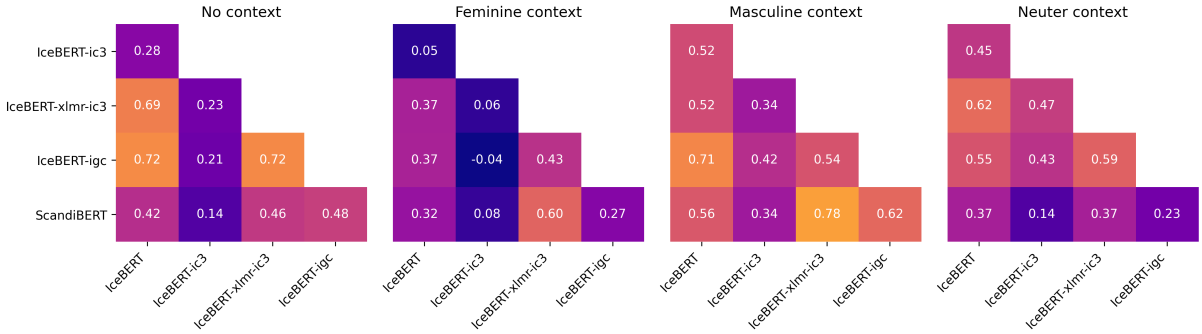

Figure 4: Correlation coefficients for feminine pronoun scores of pairs of language models with no context and context considered. A high value indicates agreement in how the models assign feminine pronoun scores to an occupation title.

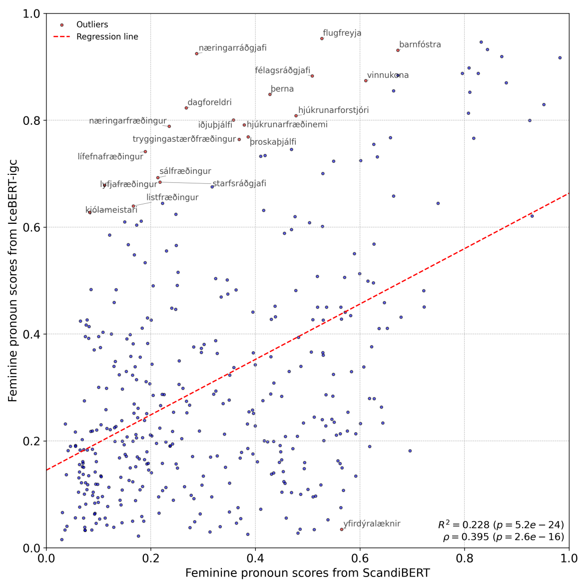

Figure 5: Comparison of the feminine pronoun scores for `ScandiBERT` and `IceBERT-igc`. Outliers that deviate by more than two standard deviations from the regression line are labeled specifically.

pants what they thought the wages of certain professions were and state their opinions on which wages would be fair for said professions without any information about the actual wages. The results show that nearly all of the participants thought the wage differences between high-paying jobs and low-paying jobs should be less than they perceived it to be but the results also indicate that most of the participants did think that a certain wage difference is desireable.

It is not clear exactly how ideas of prestige and class division affect our results. It is, however, worth noting that while Iceland has consistently placed at the top of the Global Gender Gap Index as measured by the World Economic Forum for more than a decade (World Economic Forum, 2023), the proportion of women in low-income jobs is still significantly higher than that of men (Statistics Iceland, 2023). Wheras it is certainly expected that grammatically feminine jobs are rather preceded by a feminine pronoun (notably, 'nurse' is grammatically masculine and yet almost exclusively judged to be feminine), it is possible that these 'women's jobs' are seen as having little prestige and that this influences our models' predictions.

8. Conclusion

Icelandic society, and society in general, has evolved significantly throughout the years. In the early 20th century, women typically did not participate in the workforce and any managerial position was sure to be occupied by a man. This is no longer the case, as indicated by the occupational distribution frequencies provided by Statistics Iceland. However, these societal changes might not be adequately reflected in our language models. As previously stated, the output quality of large language models is largely determined by the amount of training data they receive. Naturally, researchers tend to use all available data to train such models in order to optimize their results. But not all data will be recent. In the case of our models, the oldest parts of the training data go back as far as 1909. This means that even though society has changed, the models may not be aware of these changes. They might inherit the prejudice of the past, potentially reopening wounds that had previously been healed. Our results indicate that the biases present in these models cannot be explained by a single reason. Rather, they are a complex combination of societal, linguistic, and data selection bias as well as potential bias amplifications due to the choice of the BERT architecture.

9. Aknowledgements

Steinunn Rut Friðriksdóttir was supported by The Ludvig Storr Trust no. LSTORR2023-93030.

10. Bibliographical References

Abubakar Abid, Maheen Farooqi, and James Zou. 2021. Persistent Anti-Muslim Bias in Large Language Models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 298–306.

Jón G. Bernburg and Sigrún Ólafsdóttir. 2023. Hve mikill á launamunur að vera? Viðhorf almennings. *Vísbending*, 41(27):3–4.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. *arXiv preprint arXiv:2005.14050*.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. *Advances in Neural Information Processing Systems*, 29.

David DeFranza, Himanshu Mishra, and Arul Mishra. 2020. How Language Shapes Prejudice Against Women: An Examination Across 45 World Languages. *Journal of Personality and Social Psychology*, 119(1):7.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT*, volume 1, page 2.

Mostafa Elaraby, Ahmed Y. Tawfik, Mahmoud Khaled, Hany Hassan, and Aly Osama. 2018. Gender Aware Spoken Language Translation Applied to English-Arabic. In *2018 2nd International Conference on Natural Language and Speech Processing (ICNLSP)*, pages 1–6.

Joel Escudé Font and Marta R. Costa-jussà. 2019. Equalizing Gender Bias in Neural Machine Translation with Word Embeddings Techniques. In *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, pages 147–154, Florence, Italy. Association for Computational Linguistics.

Hagstofa Íslands. 2009. Ístarf95: Íslensk starfaflokkun með skỳringum og dæmum [Icelandic Classification of Occupations] (2nd edition).

Þóra Björk Hjartardóttir. 2004. Baráttan um orðin. orðanotkun tengd samkynhneigð. *Íslenskt mál*, 26:83–122.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. 2022. Training Compute-Optimal Large Language Models. *arXiv preprint arXiv:2203.15556*.

Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black, and Yulia Tsvetkov. 2019. Measuring Bias in Contextualized Word Representations. *arXiv preprint arXiv:1906.07337*.

Susan Leavy, Gerardine Meaney, Karen Wade, and Derek Greene. 2020. Mitigating Gender Bias in Machine Learning Data Sets. In *Bias and Social Aspects in Search and Recommendation: First International Workshop, BIAS 2020, Lisbon, Portugal, April 14, Proceedings 1*, pages 12–26. Springer.

Zihao Li, Zhuoran Yang, and Mengdi Wang. 2023. Reinforcement Learning with Human Feedback: Learning Dynamic Choices via Pessimism. *arXiv preprint arXiv:2305.18438*.

Wendy Martyna. 1980. Beyond the "He/Man" Approach: The Case for Nonsexist Language. *Signs: Journal of Women in Culture and Society*, 5(3):482–493.

Guðmundur Oddsson. 2016. Neoliberal Globalization and Heightened Perceptions of Class Division in Iceland. *The Sociological Quarterly*, 57(3):462–490.

Ágústa Þorbergsdóttir. 2021. Skỳrsla um kynhlutlaust mál. *Íslensk málnefnd*, 29:2022.

Guðrún Þórhallsdóttir. Hvað er kynhlutleysi? *Málfregnir*, page 5.

Shiva Omrani Sabbaghi and Aylin Caliskan. 2022. Measuring Gender Bias in Word Embeddings of Gendered Languages Requires Disentangling Grammatical Gender Signals. *arXiv preprint arXiv:2206.01691*.

Danielle Saunders and Bill Byrne. 2020. Reducing Gender Bias in Neural Machine Translation as a Domain Adaptation Problem. *arXiv preprint arXiv:2004.04498*.

Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. 2021. Gender Bias in Machine Translation. *Transactions of the Association for Computational Linguistics*, 9:845–874.

Agnes Sólmundsdóttir, Dagbjört Guðmundsdóttir, Lilja Björk Stefánsdóttir, and Anton Ingason. 2022. Mean Machine Translations: On Gender Bias in Icelandic Machine Translations. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3113–3121.

Artūrs Stafanovičs, Toms Bergmanis, and Mārcis Pinnis. 2020. Mitigating Gender Bias in Machine Translation with Target Gender Annotations. *arXiv preprint arXiv:2010.06203*.

Statistics Iceland. 2023. Óleiðréttur launamunur karla og kvenna 9,1% árið 2022.

Steinþór Steingrímsson, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson, and Jón Guðnason. 2018. Risamálheild: A Very Large Icelandic Text Corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, LREC 2018, Miyazaki, Japan.

Samia Touileb and Debora Nozza. 2022. Measuring Harmful Representations in Scandinavian Language Models. *arXiv preprint arXiv:2211.11678*.

Eva Vanmassenhove, Christian Hardmeier, and Andy Way. 2019. Getting Gender Right in Neural Machine Translation. *arXiv preprint arXiv:1909.05088*.

Pranav Narayanan Venkit, Mukund Srinath, and Shomir Wilson. 2022. A Study of Implicit Bias in Pretrained Language Models Against People with Disabilities. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 1324–1332.

Angelina Wang and Olga Russakovsky. 2021. Directional Bias Amplification. In *International Conference on Machine Learning*, pages 10882–10893. PMLR.

World Economic Forum. 2023. The Global Gender Gap Report 2023.

Xianjun Yang, Xiao Wang, Qi Zhang, Linda Petzold, William Yang Wang, Xun Zhao, and Dahua Lin. 2023. Shadow Alignment: The Ease of Subverting Safely-Aligned Language Models. *arXiv preprint arXiv:2310.02949*.

Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang. 2017. Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints. *arXiv preprint arXiv:1707.09457*.

Pei Zhou, Weijia Shi, Jieyu Zhao, Kuan-Hao Huang, Muhao Chen, Ryan Cotterell, and Kai-Wei Chang. 2019. Examining Gender Bias in Languages with Grammatical Gender. *arXiv preprint arXiv:1909.02224*.

11. Language Resource References

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. *Science*, 356(6334):183–186.

Debora Nozza, Federico Bianchi, Dirk Hovy, et al. 2021. HONEST: Measuring Hurtful Sentence Completion in Language Models. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics.

Vésteinn Snæbjarnarson, Haukur Barri Símonarson, Pétur Orri Ragnarsson, Svanhvít Lilja Ingólfsdóttir, Haukur Páll Jónsson, Vilhjálmur Þorsteinsson, and Hafsteinn Einarsson. 2022. A Warm Start and a Clean Crawled Corpus–A Recipe for Good Language Models. *arXiv preprint arXiv:2201.05601*.

Vésteinn Snæbjarnarson, Annika Simonsen, Goran Glavaš, and Ivan Vulić. 2023. Transfer to a Low-Resource Language via Close Relatives: The Case Study on Faroese. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, Tórshavn, Faroe Islands. Linköping University Electronic Press, Sweden.

A. Further Model Comparison

We computed Spearman's rank correlation coefficients to compare language models, which we show in Figure 6. We observed a somewhat similar structure to the one for correlation coefficients in Figure 4.

B. Highest and Lowest Scoring Jobs

We averaged the feminine pronoun scores for each job title over all the models. The results are shown in Table 4 and Table 5.

C. Appendix: Occupation Categories (English Translations)

Managerial posts: Officeholders Ombudsman of the Parliament (Umboðsmaður Alþingis) Consul General (aðalræðismaður) The President of Iceland (forseti Íslands) District magistrate (sýslumaður) Town mayor (bæjarstjóri) City mayor (borgarstjóri) Permanent Secretary of State (ráðuneytisstjóri) Official at the Municipal Council (sveitarstjórnarmaður) Office head of the Parliament (skrifstofustjóri Alþingis) Head of the Parliament (forseti Alþingis) Chairman of a political party (formaður stjórnmálaflokks) Chairman of the

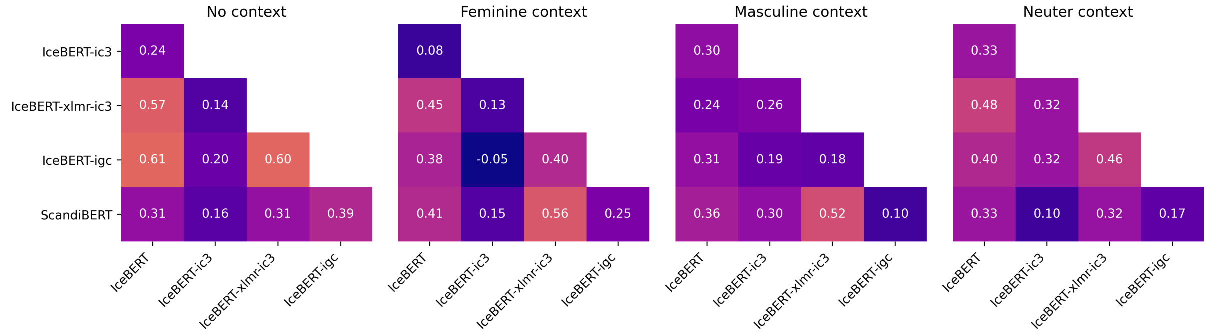

Figure 6: Spearman's rank correlation coefficients for feminine pronoun scores of pairs of language models with no context and context considered. A high value indicates agreement in how the models assign feminine pronoun scores to an occupation title.

Job Name	Score ± std
ljósmóðir *midwife*	0.852 ± 0.134
vinnukona *female laborer*	0.822 ± 0.148
nunna *nun*	0.816 ± 0.106
hjúkrunarfræðingur *nurse*	0.813 ± 0.079
flugfreyja *female flight attendant*	0.792 ± 0.154
svæfingarhjúkrunarfræðingur *anesthetic nurse*	0.768 ± 0.111
barnahjúkrunarfræðingur *pediatric nurse*	0.766 ± 0.205
þerna *maid*	0.743 ± 0.161
húshjálp *housemaid*	0.741 ± 0.234
öldrunarhjúkrunarfræðingur *geriatric nurse*	0.726 ± 0.185

Table 4: Top 10 job titles with respect to feminine pronoun scores.

Job Name	Score ± std
prófastur *archdeacon*	0.048 ± 0.044
ræstir *cleaner*	0.065 ± 0.052
skipasmiður *shipbuilder*	0.070 ± 0.038
fangelsismálastjóri *director-general of prison and probation administration*	0.073 ± 0.055
eldsmiður *blacksmith*	0.088 ± 0.056
prentsmiður *printer*	0.090 ± 0.061
aðalræðismaður *consul general*	0.094 ± 0.034
vaktmaður *watchman*	0.095 ± 0.043
bakari *baker*	0.098 ± 0.041
blikksmiður *tinsmith*	0.099 ± 0.055

Table 5: Bottom 10 job titles with respect to feminine pronoun scores.

City Council (forseti borgarstjórnar) Chairman of a District Council (oddviti) Director of a Local Council (sveitarstjóri) Congressman (alþingismaður) Town Representative (bæjarfulltrúi) Prime Minister (forsætisráðherra) Minister of the Environment (umhverfisráðherra) Minister of Education (menntamálaráðherra) Minister of Agriculture (landbúnaðarráðherra) Minister of Fisheries (sjávarútvegsráðherra) Minister of Justice (dómsmálaráðherra)

Managerial posts: CEOs and Directors University Rector (háskólarektor) Director General of Public Health (landlæknir) Radio Director (útvarpsstjóri) Savings Bank Manager (sparisjóðsstjóri) Bank Manager (bankastjóri) District Heating Manager (hitaveitustjóri) Forestry Director (skógræktarstjóri) Museum Executive Director (framkvæmdastjóri safns) National Librarian (landsbókavörður) Theatre Director (sviðsstjóri) Print Shop Manager (prentsmiðjustjóri) Human Resources Manager (starfsmannastjóri) Primary School Principal (skólastjóri grunnskóla) Executive Director (framkvæmdastjóri) Music School Principal (skólastjóri tónlistarskóla) Financial Manager (fjármálastjóri) Theatre Manager (leikhússtjóri) Program Director (dagskrárstjóri) Nursing Executive Director (hjúkrunarframkvæmdastjóri) Chief Executive Officer of Nursing (hjúkrunarforstjóri) Kindergarten Principal (leikskólastjóri) Director of Internal Revenue (ríkisskattstjóri) TV Managing Director (sjónvarpsstjóri) Fisheries Manager (útgerðarstjóri) Building Superintendent (húsameistari) Construction Manager (byggingarstjóri) Building Representative (byggingarfulltrúi) Warehouse Manager (lagerstjóri) Inventory Manager (birgðastjóri) Quality Manager (gæðastjóri) Restaurant Manager (veitingastjóri) Hotel Manager (hótelstjóri) Purchasing Manager (innkaupastjóri) Director of the Social Welfare Service (félagsmálastjóri) Operations Manager (rekstrarstjóri) Statistical Office Manager (hagstofustjóri) Director-General of Prison and Probation Administration (fangelsismálastjóri) Fire Safety Manager (brunamálastjóri) Assistant Manager (aðstoðarforstjóri) Assistant Director General of Public Health (aðstoðarlandlæknir) Assistant Trade Manager (aðstoðarverslunarstjóri) Assistant Executive Director (aðstoðarframkvæmdastjóri) Assistant Bank Manager (aðstoðarbankastjóri)

Specialized jobs: Physics, engineering and math Physicist (eðlisfræðingur) Mathematician (stærðfræðingur) Physiologist (lífeðlisfræðingur) Actuary (tryggingastærðfræðingur) Chemist (efnafræðingur) Chemical Engineer (efnaverkfræðingur) Biochemist (lífefnafræðingur) Statistician (tölfræðingur) Architect (arkitekt) Naval Architect (skipaverkfræðingur) Municipal Engineer (bæjarverkfræðingur) Landscape Architect (landslagsarkitekt) Structural Engineer (byggingarverkfræðingur) Interior Architect (innanhússarkitekt) Urban Planner (skipulagsfræðingur) Mechanical Engineer (vélaverkfræðingur) Technician (tæknifræðingur) Geophysicist (jarðeðlisfræðingur)

Specialized jobs: Natural sciences, Biology, Health sciences Oceanographer (haffræðingur) Seismologist (jarðskjálftafræðingur) Ecologist (vistfræðingur) Biologist (líffræðingur) Ornithologist (fuglafræðingur) Meteorologist (veðurfræðingur) Fisheries Scientist (sjávarútvegsfræðingur) Zoologist (dýrafræðingur) Astronomer (stjörnufræðingur) Health Inspector (heilbrigðisfulltrúi) Geographer (landfræðingur) Botanist (grasafræðingur) Glaciologist (jöklafræðingur) Marine Biologist (sjávarlíffræðingur) Naturalist (náttúrufræðingur) Nutritionist (næringarfræðingur) Nutritional Advisor (næringarráðgjafi) Speech Therapist (talmeinafræðingur) Physiotherapist (sjúkraþjálfari) Medical Masseur (sjúkranuddari) Occupational Therapist (iðjuþjálfi) Dental Technician (tannsmiður) Dentist (tannlæknir) Cardiac Surgeon (hjartaskurðlæknir) Physician (læknir) Surgeon (skurðlæknir) General Practitioner (heimilislæknir) Chief Veterinary Officer (yfirdýralæknir) Molecular Biologist (lífeindafræðingur) Pharmaceutical Technician (lyfjatæknir) Veterinarian (dýralæknir) Plastic Surgeon (lýtalæknir) Ophthalmologist (augnlæknir) Oncologist (krabbameinslæknir) Chief Physician (forstöðulæknir) Pediatrician (barnalæknir) Pharmacist (lyfjafræðingur) Medical Student (læknanemi) Gynecologist (kvensjúkdómalæknir) Nursing Student (hjúkrunarfræðinemi) Psychiatric Nurse (geðhjúkrunarfræðingur) Nurse (hjúkrunarfræðingur) Surgical Nurse (skurðhjúkrunarfræðingur) Geriatric Nurse (öldrunarhjúkrunarfræðingur) Emergency Nurse (bráðahjúkrunarfræðingur) Home Care Nurse (heimahjúkrunarfræðingur) Anesthetic Nurse (svæfingarhjúkrunarfræðingur) Pediatric Nurse (barnahjúkrunarfræðingur) General Practice Nurse (heilsugæsluhjúkrunarfræðingur) Intensive Care Nurse (gjörgæsluhjúkrunarfræðingur) Midwife (ljósmóðir)

Specialized jobs: Education and Pedagogy Professor (prófessor) Lecturer (lektor) University Teacher (háskólakennari) Adjunct (aðjúnkt) Associate Professor (dósent) High School Teacher (framhaldsskólakennari) Primary School Teacher (grunnskólakennari) Kindergarten Teacher (leikskólakennari) Developmental Therapist (þroskaþjálfi) Prevention Counsellor (forvarnarfulltrúi) Educational Representative (fræðslufulltrúi) Sports Representative (íþróttafulltrúi) Director of Studies (kennslustjóri) Academic Director (námsstjóri) Career Counselor (starfsráðgjafi) Academic Counselor (námsráðgjafi) Driving Instructor (ökukennari) School Assistant (skólaliði) Flight Instructor (flugkennari) Support Representative (stuðningsfulltrúi)

Specialized jobs: Business, Audit, Marketing Economist (hagfræðingur) Business Economist (rekstrarhagfræðingur) Industrial Organization Expert (iðnrekstrarfræðingur) Auditor (endurskoðandi) State Auditor (ríkisendurskoðandi) Advertising Director (auglýsingastjóri) PR Representative (kynningarfulltrúi) Information Officer (upplýsingafulltrúi) Marketing Director (markaðsstjóri) Marketing Consultant (markaðsráðgjafi)

Specialized jobs: Law Chief Justice of the Supreme Court (forseti hæstaréttar) Attorney General (ríkislögmaður) Prosecutor (saksóknari) State Prosecutor (ríkissaksóknari) State Mediator (ríkissáttasemjari) District Court Judge (héraðsdómari) Registrar/Court Administrator (dómstjóri) Trustee/Bankruptcy Administrator (skiptastjóri) Civil Lawyer (borgarlögmaður) Supreme Court Attorney (hæstaréttarlögmaður) District Court Attorney (héraðsdómslögmaður) Lawyer (lögfræðingur) Supreme Court Judge (hæstaréttardómari)

Specialized jobs: Social Sciences and Humanities Linguist (málvísindamaður) Historiographer/History Writer (söguritari) Historian (sagnfræðingur) Grammarian/Linguist (málfræðingur) Art Historian (listfræðingur) Anthropologist (mannfræðingur) Political Scientist (stjórnmálafræðingur) Genealogist (ættfræðingur) Theologian (guðfræðingur) Philosopher (heimspekingur) Literary Scholar (bókmenntafræðingur) Sociologist (félagsfræðingur) Folklorist (þjóðháttafræðingur) Social Counselor (félagsráðgjafi) Psychologist (sálfræðingur) Psychoanalyst (sálgreinir)

Specialized jobs: Art Sculptor (myndhöggvari) Concertmaster (konsertmeistari) Clown (trúður) Cartoonist (skopmyndateiknari) Singer (söngvari) Artist/Painter (listmálari) Actor (leikari) Prompter (hvíslari) Organist (organisti) Musician/Instrumentalist (hljóðfæraleikari) Opera Singer (óperusöngvari) Photographer (ljósmyndari) Cellist (sellóleikari) Poet (skáld) Composer (tónskáld) Film Director (kvikmyndaleikstjóri) Musician (tónlistarmaður) Playwright (leikskáld) Music Director (tónlistarstjóri) Orchestra Conductor (hljómsveitarstjóri) Exhibition Curator (sýningarstjóri) Choir Director (kórstjóri) Visual Artist (myndlistarmaður) Organ Player (orgelleikari) Flutist (flautuleikari) Pianist (píanóleikari) Choreographer (danshöfundur) Violinist (fiðluleikari) Vocal Coach (söngstjóri) Impersonator (eftirherma)

Specialized jobs: Media Publishing Editor (útgáfustjóri) News Reporter (fréttamaður) Journalist (blaðamaður) Correspondent (blaðafulltrúi) Program Producer (dagskrárgerðarmaður) News Editor (fréttastjóri) Editor-in-Chief (ritstjóri)

Specialized jobs: Religion Archdeacon (prófastur) Bishop (biskup) Monk (munkur) Priest (prestur) Nordic Heathenism High Priest (allsherjargoði) Rabbi (rabbíni) Abbot (ábóti) Hospital Chaplain (sjúkrahúsprestur) Suffragan Bishop (vígslubiskup) Missionary (trúboði) Christian Missionary (kristniboði) Church Warden (kirkjuvörður) Deacon (djákni) Assistant to a Priest (meðhjálpari) Nun (nunna)

Specialized jobs: Engineering and Science Programmer/Software Developer (forritari) Electrical Engineer (rafmagnstæknifræðingur) Mechanical Engineer (véltæknifræðingur) Sound Engineer (hljóðmaður) Cinematographer (kvikmyndatökumaður) Recording Director (upptökustjóri) Cameraman (myndatökumaður) Computer Scientist (tölvunarfræðingur) Software Engineer (hugbúnaðarverkfræðingur) Computer Engineer (reikniverkfræðingur) IT Engineer (tölvutæknifræðingur) Civil Engineer (byggingaverkfræðingur) Environmental Engineer (umhverfisverkfræðingur) Maritime Pilot (hafnsögumaður) Aviation Director (flugmálastjóri) Harbour Master (hafnarvörður) Airport Director (flugvallarstjóri) Air Traffic Controller (flugvirki) Port Director (hafnarstjóri) Flight Engineer (flugvélstjóri)

Specialized jobs: Consultants, Secretaries, Customs control Securities Broker (verðbréfasali) Auctioneer (uppboðshaldari) Real Estate Agent (fasteignasali) Financial Advisor (fjármálaráðgjafi) Bishop's Secretary (biskupsritari) Town Clerk (bæjarritari) City Clerk (borgarritari) Presidential Secretary (forsetaritari) Embassy Secretary (sendiráðsritari) Bookkeeper (bókari) Chief Bookkeeper (aðalbókari) Tax Director (skattstjóri) Customs Officer (tollvörður) Customs Director (tollstjóri) Tax Investigation Director (skattrannsóknarstjóri)

Office jobs School Secretary (skólaritari) Personal Secretary (einkaritari) Receptionist (móttökuritari) Accountant (bókhaldari) Payroll Representative (launafulltrúi) Document Custodian (skjalavörður) Librarian (bókavörður) Customer Service Representative (þjónustufulltrúi)

Service and Carer jobs Steward (framreiðslumaður) Head Waiter (yfirþjónn) Waiter (þjónn) Maid (þerna) Nanny (barnfóstra) Daycare Provider (dagforeldri) Au Pair (au-pair) Security Guard (öryggisvörður) Lifeguard (lífvörður) Night Watchman (næturvörður) Janitor (húsvörður) Swimming Pool Lifeguard (sundlaugarvörður) Detective (rannsóknarlögreglumaður) Police Chief (lögregluvarðstjóri) Watchman (vaktmaður) Supervisor (eftirlitsmaður) Prison Guard (fangavörður) Firefighter (slökkviliðsmaður) Rescue Worker (björgunarsveitarmaður) Paramedic (sjúkraflutningamaður) Fishing Warden (veiðivörður) Forest Ranger (skógarvörður)

Artisans Master House Painter (málarameistari) Barber (rakari) Hairdresser (hárskeri) Beauty Stylist (snyrtir) Hairstylist (hárgreiðslumaður) Painter (málari) Beautician (snyrtifræðingur) Master Hairstylist (hárgreiðslumeistari) Builder (húsasmiður) Master Carpenter (trésmíðameistari) Furniture Maker (húsgagnasmiður) Electrician (rafvirki) Master Electrician (rafvirkjameistari) Bookbinder (bókbindari) Printer (prentsmiður) House Painter (húsamálari) Stone Mason (steinsmiður) Shipbuilder (skipasmiður) Blacksmith (eldsmiður) Tin Smith (blikksmiður) Boat Builder (bátasmiður) Car Mechanic (bifreiðasmiður) Goldsmith (gullsmiður) Ironsmith (járnsmiður) Watchmaker (úrsmiður) Wood Turner (rennismiður) Mechanic (vélvirki) Automobile Mechanic (bifvélavirki) Brewer (bruggari) Master Chef (matreiðslumeistari) Food Scientist (matvælafræðingur) Cook (matráður) Baker (bakari) Butcher (slátrari) Chef (kokkur) Head Chef (yfirkokkur) Dairy Scientist (mjólkurfræðingur) Meat Industry Worker (kjötiðnaðarmaður) Fish Inspector (fiskmatsmaður) Fish Scientist (fiskifræðingur) Net Maker (netagerðarmaður) Shoemaker (skósmiður) Weaver (vefari) Furrier (feldskeri) Tailor (klæðskeri) Dressmaker (kjólameistari) Fashion Designer (tískuhönnuður) Costume Designer (búningahönnuður) Clothes Designer (fatahönnuður)

Motorists School Bus Driver (skólabílstjóri) Milk Truck Driver (mjólkurbílstjóri) Pilot/Aviator (flugmaður) Airline Pilot (flugstjóri) Personal Driver (einkabílstjóri) Delivery Truck Driver (sendibílstjóri) Freight Truck Driver (vörubílstjóri) Small Boat Fisherman (trillukarl) Taxi Driver (leigubílstjóri) Ship Captain (skipherra)

Non-specialist jobs: Sales and Services Shopkeeper (búðarmaður) Bartender (barþjónn) Fishmonger (fisksali) Newspaper Carrier (blaðberi) Telephone Operator (símavörður) Courier (sendill) Doorman/Gatekeeper (dyravörður) Weather Observer (veðurathugunarmaður) Lighthouse Keeper (vitavörður) Flight Attendant (Male) (flugþjónn) Flight Attendant (Female) (flugfreyja) Housemaid (húshjálp) Laborer (Female) (vinnukona) Laborer (Male) (vinnumaður) Drug Salesperson (lyfsali) Pharmacist (apótekari) Mail Carrier (bréfberi) Cleaner (ræstir) Cleaning Technician (ræstitæknir) Mower (sláttumaður) Horse Farmer (hrossabóndi) Shepherd (smali) Sheep Farmer (sauðfjárbóndi) Cattle Farmer (kúabóndi) Potato Farmer (kartöflubóndi) Horse Trainer (tamningamaður) Gardening Supervisor (garðyrkjustjóri) Horticulturist (garðyrkjufræðingur) Vegetable Farmer (garðyrkjubóndi)