

Universidade de São Paulo  
Instituto de Matemática e Estatística  
Bacharelado em Ciência da Computação

Ludmila Ferreira Vicente e Silva

**Inovação de Gênero:  
Detecção de padrões de viés de gênero em textos de  
websites**

São Paulo  
Novembro de 2018

**Inovação de Gênero:**  
**Detecção de padrões de viés de gênero em textos de**  
**websites**

Monografia final da disciplina  
MAC0499 – Trabalho de Formatura Supervisionado.

Supervisor: Prof. Dr. Alfredo Goldman vel Lejbman  
[ Cosupervisora: Profa. Dra. Cláudia de Oliveira Melo ]

São Paulo  
Fevereiro de 2019

# Agradecimentos

À minha família, Hélio, Nilza, Nadiajda, Vovó DoCarmo e família, pois graças a todos os seus esforços, segurança e amor pude seguir meu caminho e me tornar quem sou hoje. Ao Ian pelo companherismo, carinho e dedicação durante todo os incríveis momentos ao meu lado. Aos meu queridos e imprescindíveis amigos Gervásio, Victor, Matheus, Vinicius, Gustavo, Luciana, Milene, Helena, João, Maurício, Renato, Luiz joaquim, Áurea, Pan Marcondes e todos que não pude citar aqui, que me ajudaram durante a graduação de inúmeras maneiras e tornaram meus anos no IME em uma vivência de muito aprendizado e amizade. Ao meu orientador, Alfredo Goldman, pela paciência, compreensão, oportunidades, por acreditar tanto em mim em momentos que nem eu mesma acreditava e principalmente pelo bom humor. A minha co-orientadora Cláudia Melo por me apresentar a Inovação de Gênero. E, por último, às minhas gatinhas Ada, Minerva e Selina por terem sido companhias tão doces durante os últimos anos.

Todos vocês foram imensuravelmente importantes durante todos esses anos e eu jamais serei capaz de conceber em palavras minha gratidão.

Muito obrigada.



# Resumo

Nesse trabalho, aplicamos Inovação de Gênero e processamento de linguagem natural para a Correção de Conhecimento em textos retirados de websites informativos. Com objetivo de detectar algum viés de gênero em textos que transmitem informação, extraímos um corpus de colunas da internet. Esses textos foram parseados, anotados e classificados por aprendizado de máquina a partir de rótulos de gênero. A análise sugere uma ligação tênue entre determinados termos e gênero feminino.

**Palavras-chave:** inovação de gênero, correção de conhecimento, processamento de linguagem natural.



# Abstract

In this work, we have applied Gender Innovation and Natural Language Processing to the Correction of Knowledge in texts retrieved from news and general interests websites. Aiming to detect gender bias in informative texts, we extracted a corpus of blog columns and open editorial from the internet. This texts were be parsed, annotated and classified using machine learning technology into gender labels. The analysis sugests sligth connections between certains terms and the female gender.

**Keywords:** gendered innovations, fixing the knowledge, natural language processing, keyword3.



# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Desenvolvimentos</b>	<b>3</b>
2.1	Análise Teórica . . . . .	4
2.1.1	O papel da mulher na sociedade . . . . .	4
2.1.2	Inovação de Gênero . . . . .	7
2.1.3	Aspectos da Linguagem . . . . .	9
2.1.4	Processamento de Linguagem Natural . . . . .	10
2.1.5	Objetivo . . . . .	11
2.1.6	Estudo de Caso . . . . .	12
2.1.7	Ferramentas . . . . .	13
2.2	Desenvolvimento Prático . . . . .	14
<b>3</b>	<b>Conclusões</b>	<b>19</b>
	<b>Referências Bibliográficas</b>	<b>21</b>



# Capítulo 1

## Introdução

Nas ciências exatas existe uma grande diferença entre o número de mulheres e homens e, como a ciência foi historicamente desenvolvidas por homens, a cultura do meio se torna predominantemente masculina. Ambientes que eram majoritariamente ocupados por mulheres no início do seu desenvolvimento, como a Ciência da Computação, sofreram uma vertiginosa inversão em relação ao número em gênero, dado que homens foram cada vez mais atraídos para área[23]. Este trabalho levanta o questionamento sobre os motivos que levam a mulher se afastar de áreas de ciências exatas.

Algumas medidas nas esferas de Gênero e Educação são nosso ponto de partida. Entidades internacionais propuseram soluções para que pesquisa e educação se tornem meios mais produtivos e igualitários como descrito no relatório *Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation* da [Comissão Europeia \(2012\)](#) que indica que um dos aspectos a ser aplicado para que a equidade de gênero seja atingida é a incorporação de Análise de Gênero na Ciência, sendo essa uma de nossas motivações. Além disso, a [Organização das Nações Unidas \(2015\)](#) criou um plano de ações que visa o Desenvolvimento Sustentável do planeta com uma série de objetivos a serem alcançados até 2030. Dois destes objetivos também estão nas motivações do nosso trabalho: “Garantir uma educação de qualidade inclusiva e equitativa e promover oportunidades de aprendizagem ao longo da vida para todos”, e “Alcançar a igualdade de gênero e empoderar todas as mulheres e meninas”.

Refletimos sobre alguns pontos como o papel da mulher na sociedade, as condições em que ser social se desenvolve, a língua e a cultura, que refletem o ambiente social, e como todas essas condições podem influenciar na formação do ser. Existem iniciativas para compreender o que leva a essa disparidade e descobrir quais medidas podem ser tomadas para que a produção científica tenha um ambiente mais diversificado, igualitário e produtivo. Essas iniciativas são chamadas de Inovação de Gênero. Nosso estudo é motivado pela ideia de que diversificar o meio e a maneira que a ciência é feita é uma forma de inserir novas abordagens para os problemas estudados, o que pode contribuir para trazer novas perspectivas na pesquisa científica.

Nossa iniciativa para promover a Inovação de Gênero é uma análise do viés que pode estar atrelado ao gênero, obtida a partir de processamento de linguagem natural de textos obtidos em websites jornalísticos voltados para um público diverso, que mostrou alguns traços que podem indicar relações entre determinados termos, temas e o gênero.



# Capítulo 2

## Desenvolvimentos

Apesar do papel expressivo de várias mulheres na ciência, como Ada Lovelace[37], Marie Curie[35] e Grace Hopper[36], o número delas vem diminuindo em proporção ao número de homens na área. É importante compreender os motivos que levaram a essa diminuição e também propôr soluções para tornar a área equalitária.

Pesquisamos sobre a condição da mulher nos meios sociais e como a cultura ocidental endossa essa condição criando estereótipos que ditam qual deve ser o comportamento da mulher. A análise de como a língua é formada e transmitida no permite compreender que uma vez que a cultura e a linguagem estão profundamente interligadas em seu desenvolvimento, esterótipos ligados a imagem da mulher na sociedade podem ser refletidos na linguagem perpetuando o estigma relacionado ao gênero feminino. Dessa maneira pudemos embasar a hipótese de que pode existir um viés de gênero no discurso. Uma vez que o discurso tem como objetivo transmitir alguma espécie de informação que é dotada de sentido para o locutor, o preconceito em relação ao gênero feminino pode ser carregado nessa mensagem e informar de maneira enviesada.

Nos questionamos como podemos encontrar padrões em textos que indiquem problemas referentes a viés de gênero e que tornem o meio em que o texto está imerso em um espaço menos inclusivo. Nosso foco é a Correção de Conhecimento: identificar e corrigir a maneira com que informação é transmitida, para isso vamos analisar a forma que o gênero feminino é representado em textos da Língua Portuguesa presentes em websites.

O desenvolvimento prático foi feito na linguagem de programação *Python* e envolveu extrair dados de notícias de websites a partir do webcrawler *Scrapy*, além buscar alguma informação sobre possíveis vieses nesses dados através de classificadores presentes na biblioteca *NLTK*. Com esses passos chegamos às conclusões do nosso estudo.

## 2.1 Análise Teórica

### 2.1.1 O papel da mulher na sociedade

Por que o número de mulheres nas áreas de ciências exatas é tão discrepante em relação ao número de homens? Esse questionamento nos remete a um grande debate sobre qual papel é atribuído a mulher na sociedade como um todo.

O primeiro ponto relevante a ser mencionado é como a mulher era enxergada no contexto social através da história. Como descrito por [Butler \(2003\)](#) em seu livro sobre questões de gênero:

Em primeiro lugar, as concepções humanistas do sujeito tendem a presumir uma pessoa substantiva, portadora de vários atributos essenciais e não essenciais. A posição feminista humanista compreenderia o gênero como um atributo da pessoa, caracterizada essencialmente como uma substância ou um “núcleo” de gênero preestabelecido, denominado pessoa, denotar uma capacidade universal de razão, moral, deliberação moral ou linguagem. Como ponto de partida de uma teoria social do gênero, entretanto, a concepção universal da pessoa é deslocada pelas posições históricas ou antropológicas que compreendem o gênero como uma relação entre sujeitos socialmente constituídos, em contextos especificáveis. Este ponto de vista relacional ou contextual sugere que o que a pessoa “é” - e a rigor, o que o gênero “é” - refere-se sempre às relações construídas em que ela é determinada. Como fenômeno inconstante e contextual, o gênero não denota um ser substantivo, mas um ponto relativo de convergência entre conjuntos específicos de relações, cultural e historicamente convergentes.

às mulheres eram destinadas tarefas domésticas e, uma vez que são responsáveis pela gestação dos filhos também eram responsáveis pela educação destes. Esse papel limitou as mulheres a um papel coadjuvante em uma sociedade em que apenas homens eram julgados aptos para tarefas tidas como mais importantes como trabalho fora de casa para prover a família enquanto as mulheres não podia exercê-las. Além disso esse contexto, em que a mulher é tida como inapta para exercer outro tipo de ocupação, propicia a formação do estigma relacionado ao gênero feminino. O estigma é explicado por [Goffman \(1975\)](#) no trecho seguinte:

A sociedade estabelece os meios de categorizar as pessoas e o total de atributos considerados como comuns e naturais para os membros de cada uma dessas categorias: Os ambientes sociais estabelecem as categorias de pessoas que têm probabilidade de serem neles encontradas. As rotinas de relação social em ambientes estabelecidos nos permitem um relacionamento com “outras pessoas” previstas sem atenção ou reflexão particular. Então, quando um estranho nos é apresentado, os primeiros aspectos nos permitem prever a sua categoria e os seus atributos, a sua “identidade social” - para usar um termo melhor do que “status social”, já que nele se incluem atributos como “honestidade”, da mesma forma que atributos estruturais, como “ocupação”.

O estigma da mulher pode surgir no momento em que ela tenta se colocar em um ambiente que a sociedade não espera que ela ocupe pois existe uma dúvida sobre a sua capacidade de exercer algum papel fora do conjunto de tarefas comumente atribuídas a ela. Além de carregar o estereótipo de dona de casa, a mulher é associada a características ditas femininas

como por exemplo: sensibilidade e subserviência que remetem a um ser frágil que pode ser controlável, enquanto homens são associados a características geralmente opostas como agressividade e liderança, que remetem a um ser independente e mais capaz de exercer tarefas relacionadas a essas características.

Assim, o papel e estigma cultural da mulher leva a uma educação diferente para crianças do sexo masculino e do sexo feminino. Garotos são estimulados desde pequenos às tarefas que são julgadas como masculinas, que estimulam o raciocínio lógico. Quebra-cabeças, videogames e esportes que exigem mais do físico são apresentados como atividades em que garotos tem uma tendência nata enquanto resta às garotas, por exemplo, bonecas, itens relacionados ao ambiente doméstico e kits de enfermeira que também reforçam o estereótipo relacionado às tarefas que deve exercer[5]. O trecho a seguir do artigo de [Sayão \(2003\)](#) explicita o resultado das brincadeiras que indicam as posições de gênero:

Em especial, nas brincadeiras que inventam, meninos e meninas demonstram que os papéis de gênero delineados muito cedo, embora na infância seja bastante possível transgredilos. Essa característica de transgressão parece ser uma manifestação típica de um momento da vida e logo será abrandada em face das convenções sociais pautadas por uma visão da ciência e por determinações sociais que normatiza lugares comportamentos e formas de ser específicos para meninos e meninas, homens e mulheres.

A brincadeira na infância faz parte do desenvolvimento e processo de aprendizado da criança. A partir dos primeiros contatos com aspectos da sociedade abordados nas brincadeiras, a criança absorve esses aspectos e começa a moldar sua visão da sociedade. Se as brincadeiras são delimitadas por gênero desde a infância, a criança aprende a distinguir os papéis sociais por gênero muitas vezes, aumentando o estigma relacionado ao gênero feminino. Dessa maneira, as pessoas são sujeitas a uma construção social desde cedo, que dita quais tarefas são apropriadas para cada gênero. Se é ensinado para garotas desde cedo que elas não se encaixam em determinado meio, as chances de dela se interessar e de fato ingressar em uma carreira considerada masculina diminui consideravelmente. O mesmo acontece no caso dos garotos, que podem ser aconselhados a não seguir determinada carreira considerada feminina.

Porém, gênero não define a capacidade ou aptidão de uma pessoa[18], é necessário estimular o contato das crianças com brinquedos, livros e conselhos que apresentem os papéis que uma pessoa pode adotar nem uma sociedade independente de gênero. Dessa maneira acreditamos que influenciadas desde crianças, mais mulheres serão incentivadas a seguir em carreiras onde o número delas é menor que o número de homens.

Além de todos os desafios enfrentados pelas mulheres desde de crianças, quando finalmente entram em um meio considerado masculino, os entraves continuam, se tornando mais uma barreira para o desenvolvimento profissional. Mulheres podem encontrar problemas no momento de avaliações para o ingresso na carreira acadêmica, como aconteceu na Universidade de Tokyo:

A Universidade de Medicina de Tóquio, no Japão, admitiu, nesta terça-feira (07/08/2018), que baixou as notas das mulheres no concurso de admissão com o objetivo de limitar o número de estudantes do sexo feminino e se desculpou por essas velhas práticas discriminatórias[21].

No início da carreira profissional encontramos mais barreiras. A startup [Textio \(2019\)](#) analisou um banco de dados de textos de descrição de vagas de emprego reais e chegou

a conclusão de que muitos dos termos usados para ilustrar o trabalho são expressões que remetem a um universo masculino. Essa referência a um ambiente em que características masculinas são primordiais pode gerar desconforto e sensação de não alinhamento em alguém do sexo feminino ao ponderar se candidatar para tais vagas.

And these characteristically masculine terms are used even more frequently in postings for higher-paying positions, according to new data from the company, which studies millions of real job listings and related information to find patterns and trends that help companies hire more efficiently - and without bias. The findings add another layer of explanation for the lack of women at the top in the business world. “It is not super surprising that the listings skew more male, when you look at composition of executive teams”, Kieran Snyder, Textio’s CEO and cofounder told The Huffington Post. Majority-male executives are probably driving the job requirements for these roles and using characteristic language.

Segundo [da Silva e Ribeiro \(2014\)](#) um aspecto relacionado ao baixo número de mulheres no meio científico é que a ciência também é uma construção social e histórica, um produto de relações de poder, logo as construções científicas não são universais, dado que existe um viés androcêntrico na ciência que entende o homem branco, heterossexual, capitalista ocidental como sujeito do conhecimento e a forma de socialização nesse meio se baseia no estereótipo de tal sujeito.

Apesar de tantos indicativos existe uma tendência a negar o desequilíbrio entre os gêneros na ciência, uma vez que ciência é um campo de exatas, é esperado que os resultados como pesquisador sejam mais relevantes do que o sexo do indivíduo. Porém mulheres podem ser expostas, por exemplo, a microagressões[38] e juízos de valores relacionados ao estigma da mulher que atrapalham o desenvolvimento da sua carreira na ciência, como exposto por [Marilia Sá Carvalho \(2015\)](#) e as outras autoras no artigo *Mulheres no mundo da ciência e da publicação científica*:

Mais sutil é o preconceito no dia a dia, que temos tendência a negar, assumindo que o gênero não deveria ter qualquer influência na avaliação. Em recente revisão sobre o viés de gênero nas publicações científicas, verificou-se sub-representação das mulheres não só entre autores, mas principalmente entre revisores e editores[13]. O viés não é uniforme: na área de matemática as mulheres são somente 15% dos pesquisadores, sendo ainda menos representadas na editoria, apenas 10%[34]. A situação é ainda mais grave quando se analisa revistas com maior prestígio acadêmico como a Science. Examinando o primeiro e o último autores de amostra dos artigos publicados em 2015, verificou-se que a proporção de mulheres, seja como autor júnior ou sênior, era um terço menor do que a sua participação nas instituições acadêmicas norte-americanas[14].

Apesar de todo esse panorama, sempre existiram mulheres que desafiaram o que era tido como padrão adentrando em tais campos ditos “masculinos”, questionando suas estruturas e reafirmando sua presença. Alterar esse modelo cultural baseado em dicotomia de gênero e torná-lo mais inclusivo é um desafio cada vez mais saliente e nesse ponto a Inovação de Gênero entra em foco.

### 2.1.2 Inovação de Gênero

Nas últimas três décadas as mulheres conseguiram aumentar seu espaço na ciência, contribuindo para o crescimento da pesquisa e trazendo novos pontos de vista a serem considerados. Elas vem ocupando cada vez mais espaço em cursos de ciências exatas, apesar de seu número em cargos altos na academia e no mercado de trabalho continuar deveras inferior ao número de homens[40]. Nesse contexto instituições e pesquisadores há muito tempo apontam tais problemas na participação acadêmica do gênero feminino, como pesquisadoras e também como uma dimensão a ser analisada na pesquisa científica[5][12][1][29].

A Inovação de Gênero tomou maior proporção na Universidade de Standford[30] em 2009 e posteriormente, em 2011 o *Directorate-General for Research & Innovation of the European Commission* criou o grupo de *Innovation through Gender* com o propósito de “prover a cientistas e engenheiros métodos práticos para a análise de sexo e gênero e desenvolver casos de estudo como ilustrações concretas de como a análise de sexo e gênero leva a novas ideias e excelência em pesquisa”. O grupo publicou o relatório *Gendered Innovations How Gender Analysis Contributes to Research*[33] que propõe alguns casos de estudo acerca de gênero na pesquisa científica dá ênfase que não consideração do gênero na pesquisa científica pode acarretar em uma perda importante de informações que poderiam levar a um melhor resultado, salienta:

Thirty years of research have revealed that sex and gender bias is socially harmful and expensive. Gender bias also leads to missed market opportunities. In engineering, for example, assuming a male default can produce errors in machine translation. In basic research, failing to use appropriate samples of male and female cells, tissues, and animals yields faulty results. In medicine, not recognizing osteoporosis as a male disease delays diagnosis and treatment in men. In city planning, not collecting data on caregiving work leads to inefficient transportation systems. It is crucially important to identify gender bias and understand how it operates in science and technology. But analysis cannot stop there. Gendered Innovations offer sophisticated methods of sex and gender analysis to scientists and engineers. Integrating these methods into basic and applied research produces excellence in science, health & medicine, and engineering research, policy, and practice.

O trecho de Londa Schiebinger (2008), uma das professoras de Stanford à frente do projeto, pode resumir a Inovação de Gênero:

By gendered innovations I mean changes in the personnel, cultures, and content of science and engineering brought about by efforts to bring more women into these fields and to reform practices that have disadvantaged women.

O relatório se baseia nas três frentes da Inovação de Gênero para analisar e propor soluções inserir o gênero na pesquisa e no mercado de trabalho a seguir.

#### **Corrigir o número de mulheres para aumentar a participação delas**

É necessário analisar de uma perspectiva história e sociológica a participação da mulher na ciência e pesquisa, e identificar os motivos pelos quais mulheres que tendem a se afastar destes ambientes, quais estímulos mulheres recebem quando escolhem tais carreias, qual é a dinâmica das instituições, quais os grandes feitos por mulheres, quais as são as grandes mulheres cientistas e de outras áreas que se destacam (dado que a ciência e história falham

em mostrá-las em comparação aos homens). A representatividade e encorajamento da mulher a seguir tal caminho são importantes passos para aumentar o número de mulheres nas áreas de exatas.

### **Corrigir as Instituições para promover igualdade de gênero nas carreiras através de mudanças estruturais em entidades de pesquisa**

A cultura é formada por suposições e valores de seus indivíduos, uma vez que crescemos em uma cultura que em que a mulher carrega um estigma, este pode estar inconscientemente atrelado ao imaginário das pessoas. São importantes os projetos para estimular a entrada e proporcionar a permanência de mulheres em áreas predominantemente ocupadas por homens, levantar o debate sobre igualdade e promover um ambiente menos hostil são fatores que podem aumentar a produtividade do meio[20][27]. As instituições devem promover projetos com objetivo de derrubar os estereótipos de gênero, como a criação de debates sobre igualdade, foco na inclusão da mulher seja na academia ou mercado de trabalho, políticas de igualdade salarial, licença maternidade e creches que são muito importantes para o desenvolvimento da criança assim como o desenvolvimento da carreira dos responsáveis. Assim cria-se um ambiente mais favorável a mulher e conseqüentemente, aumenta-se a sua plena participação em tais áreas.

### **Corrigir o conhecimento para estimular a excelência científica a partir da introdução da dimensão de gênero e sexo na pesquisa**

A exclusão de mulheres em determinados meios como a academia é algo conhecido, porém é importante levantar o debate de quais são os efeitos dessa exclusão em uma esfera maior, a do conhecimento humano. Inserir gênero como um dos focos na pesquisa, traz a necessidade de corrigir como o conhecimento é gerado e transmitido, e quebrar as estruturas arcaicas que foram incorporadas pela ciências, dado que durante muito tempo ela foi guiada em sua maioria por homens. Assim podemos abrir mais espaço para a criatividade e novas questões no futuro da pesquisa científica.

Baseado nesses três tópicos, o relatório *Gendered Innovations How Gender Analysis Contributes to Research*[33], anteriormente citado, sugere alguns estudos de caso acerca de gênero que são o ponto de partida para a incorporação da esfera de gênero na pesquisa científica.

Este trabalho tem como premissa o estudo de caso *Communicating Science Textbooks: Rethinking Language and Visual Representations*[33] (página 18) que propõe que a informação seja transmitida de maneira não-enviesada em publicações. Dado que o objetivo de uma publicação, seja ela um livro, um artigo ou texto de jornal, tem como objetivo informar e ensinar sobre o trabalho científico, deve existir um cuidado com a linguagem, uma vez que esta pode carregar viéses baseados em estereótipos relacionados ao gênero feminino. Quando se agrega uma imagem estigmatizada ao gênero em uam fonte de informação, esta pode reproduzir e reforçar preconceitos presentes no ambiente social e, uma vez que a ciência é entendida como imparcial, tal estigma pode ser absorvido pelo interlocutor como algo natural e justificado.

Neste trabalho adicionaremos a dimensão de gênero na análise de publicações textos da internet. Em específico, textos de websites em língua portuguesa voltados para a informação como blogs e jornais, que trazem informações a uma grande quantidade de pessoas. A língua portuguesa é tida como complexa O acesso à informação deve ser democratizado, qualquer

indivíduo deve possuir o direito de acessar informação de qualidade. Com essa motivação, extrapolamos para além do mercado de trabalho e academia com o objetivo de trazer a correção de conhecimento também para ambientes que tem um público mais diverso, que também pode ser afetado pelos possíveis vieses.

### 2.1.3 Aspectos da Linguagem

Para embasar nosso pressuposto de que a linguagem pode influenciar na formação do estigma e no valor de juízo que o indivíduo agrega ao gênero, recorreremos à linguística.

#### Linguagem, cultura e identidade

Como a cultura e a língua influenciam no processo de formação da identidade do indivíduo? A cultura é um conjunto complexo que inclui aspectos como crenças, saberes, modos e direitos que são desenvolvidos e adquiridos por indivíduos vivendo em conjunto. Ela é um elemento que distingue esse grupo, sendo assim um determinante de identidade. Enquanto a língua é um traço cultural herdado socialmente por indivíduos que integram um mesmo meio e é constantemente moldada e condicionada pela realidade de tal meio[8]. A língua é maneira como seres sociais se comunicam, seja ela por meio de símbolos, escrita ou falada.

A linguística nos diz, segundo [J. Mattoso Câmara Jr \(1955\)](#), que essa comunicação é um intercâmbio cultural feito a partir de um conjunto de símbolos, articulação de segmentos vocais e regras que juntos formam a linguagem. Câmara também argumenta que a língua é parte da cultura e que até certo ponto uma pode explicar a outra e como a cultura pode representar a estrutura social pois “é uma estrutura cultural modelo, que nos permite ver a estrutura social não nítida, imanente em outros aspectos da cultura”. Então aspectos da cultura podem ser representados pela língua, como cita na dissertação de [Cunha \(2012\)](#) sobre a obra do linguísta Sapir.

"Sapir afirma que a “língua não existe separada da cultura, isto é, do conjunto socialmente herdado de práticas que determina a textura de nossas vidas” [24]. Aqui, pode-se retomar a noção da língua como um fato cultural, pois a seu ver a língua é uma prática herdada socialmente inerente ao nosso cotidiano. A cultura é social, e antecedente ao indivíduo, ente este que a herda, ou seja, adquire-a junto aos demais membros de sua comunidade. Ela também é determinística, em alguma medida, pois nos fornece o conjunto de práticas que seguiremos ao longo de nossa vida: alimentação, vestuário, língua, religião, e outros tantos aspectos da vida humana, que são adquiridos socialmente. Por fim, a cultura é tudo aquilo que um ser humano faz e aprende a fazer."

A língua pode representar a cultura do meio, dado que é desenvolvida baseada nesse ambiente. Em uma cultura que várias vezes limitou o papel da mulher a um estereótipo de femininidade e trabalho doméstico, é esperado que esse preconceito seja refletido na linguagem que, como determinante de identidade, sugere aos indivíduos, qual o papel esperado para tal gênero exercer no meio social.

Se a linguagem permite-nos fazer leituras distintas e particulares do mundo e seu entorno, isso implica dizer que a forma com a qual lidamos e interagimos em sociedade influencia nas manifestações linguísticas e conseqüentemente culturais. Posto ser a partir da manifestação cultural em circunstâncias coletivas que nos autoindentificamos. [8]

Assim, entendemos que a linguagem pode influenciar nas concepções de cada indivíduo e que estigmas e estereótipos podem ser reiterados durante a comunicação, seja por qual meio ela acontecer.

### Sexismo na linguagem

Quando pensamos na língua Portuguesa, um dos vieses do meio social é refletido na linguagem no que é chamado sexismo linguístico como apontado por [Bueno \(2015\)](#) em:

"O chamado sexismo linguístico é uma forma de discriminação revelada, por exemplo quando ao referir-se a um sujeito composto em uma oração gramatical, a norma culta da língua portuguesa obriga o gênero feminino a embutir-se ao termo masculino. O sexismo na linguagem revela-se também, através de “expressões impregnadas de estereótipos, desigualdades, desrespeito, inverdades científicas, preconceito, no que diz respeito a mulheres e homens”. Por que ainda utiliza-se palavras como Homem para designar toda a espécie humana ao invés de Humanidade? Ou quais as implicações para se ter um único gênero representando lexicamente a dignidade da espécie?"

O sexismo linguístico é um reflexo das relações entre gênero na sociedade. Por que usar o artigo feminino ao se dirigir a um grupo, quando em determinada época a mulher era aconselhada a não falar sem a permissão do marido em guias de comportamento[? ]? Isso expõe mais um ponto da exclusão da mulher e a subordinação à qual a sua imagem está atrelada, mudar essa imagem errônea é um trabalho que pode ser feito dando maior atenção ao que é e como é transmitido no discurso.

É importante velar para que práticas como sexismo linguístico não passem despecebidas e sejam corrigidas, pois uma vez que a linguagem é um espelho da realidade, a representação do gênero tem que ser feita de maneira construtiva e esclarecedora pois é parte da identidade do indivíduo. De acordo com [Larrosa \(2002\)](#):

Quando fazemos coisas com as palavras, do que se trata é de como damos sentido ao que somos e ao que nos acontece, de como correlacionamos as palavras e as coisas, de como nomeamos o que vemos ou o que sentimos e de como vemos ou sentimos o que nomeamos. Nomear o que fazemos, em educação ou em qualquer outro lugar, como técnica aplicada, como práxis reflexiva ou como experiência dotada de sentido, não é somente uma questão terminológica. As palavras com que nomeamos o que somos, o que fazemos, o que pensamos, o que percebemos ou o que sentimos são mais do que simplesmente palavras. E, por isso, as lutas pelas palavras, pelo significado e pelo controle das palavras, pela imposição de certas palavras e pelo silenciamento ou desativação de outras palavras são lutas em que se joga algo mais do que simplesmente palavras, algo mais que somente palavras.

#### 2.1.4 Processamento de Linguagem Natural

A área de Processamento de Linguagem Natural é uma área da Inteligência Artificial que nos permite analisar as línguas humanas no contexto computacional. No PNL a estrutura

da língua é analisada por aprendizado de máquina a partir da segmentação dos textos em grupos essenciais (como sentenças e tokens) e do estudo das relações entre esses grupos que geram a língua. O aprendizado de máquina por ser feito de maneiras:

- **Aprendizado supervisionado:** é fornecida uma base de treinamento ao algoritmo com entradas e suas respectivas saídas a fim de encontrar uma estrutura que leva à saída desejada
- **Aprendizado não supervisionado:** O algoritmo analisa os dados sozinho a fim de encontrar padrões por si só
- **Aprendizado por reforço:** O algoritmo interage com o ambiente e recebe um feedback sobre quão “boa” é a sua ação

Optamos por utilizar uma abordagem de PNL com aprendizado supervisionado, onde nossas entradas com as respectivas saídas desejadas serão extraídas de um corpora a partir de PNL.

### 2.1.5 Objetivo

Estereótipos moldam conjuntos de características que são relacionadas também ao comportamento. A linguagem por si está permeada de vieses relacionados a estes estereótipos presentes no meio cultural e é um determinante de identidade, dessa maneira ela pode influenciar no comportamento do indivíduo no seu ambiente social. Corrigir o conhecimento envolve mudar concepções pré-estabelecidas sobre o gênero na informação, como as inverdades de que mulheres são “naturalmente” mais voltadas para ciências humanas do que exatas, ou que se interessam “naturalmente” mais por maquiagem do que por esportes. Essas tendências existem devido a construção social a qual meninas e meninos são submetidos desde a infância.

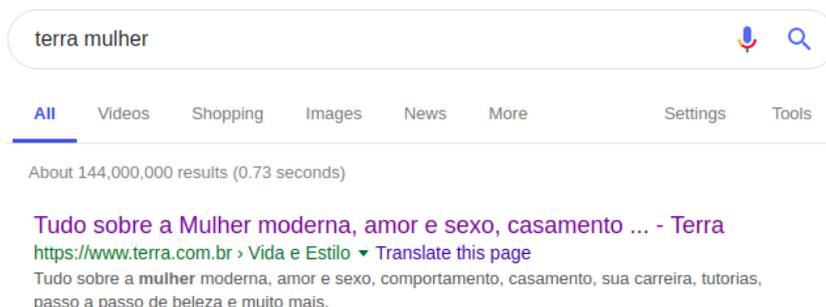
Levantamos a hipótese de que esses vieses de gênero podem ser encontrados na linguagem escrita a partir de padrões associados aos termos utilizados na linguagem e vamos testá-la com apoio de Processamento de Linguagem Natural.

### 2.1.6 Estudo de Caso

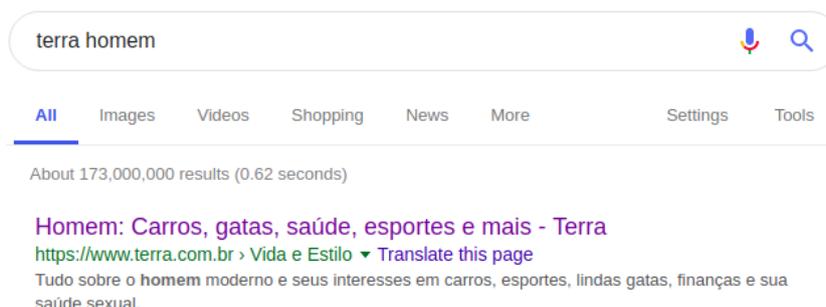
Para o nosso estudo de caso, procuramos por veículos informativos que atigissem um grande número de pessoas. Em um primeiro momento, buscamos por notícias jornalísticas acerca de dois indivíduos, um de gênero feminino e o outro masculino. Essa tentativa se mostrou falha os seguintes pontos:

- O textos foram extraídos de notícias de jornais veiculas por meio eletrônico. Como na construção do corpora não foi levado em conta que o gênero literário jornalístico busca a imparcialidade, é possível que os dados extraídos não resultassem em uma informação expressiva sobre viés de gênero.
- A análise de textos baseada em apenas um indivíduo de cada gênero pode identificar vieses que são relacionados a cada um dos sujeitos como indivíduos e não necessariamente vieses associado ao gênero.
- A base de treinamento criada a partir destes textos era claramente enviesada, pelos motivos do item anterior e dada a maneira que foi construída. A construção não envolveu um número considerável de pessoas que deveriam classificar as sentenças do texto, nem a análise estatística para atribuir um valor às sentenças que formariam a base de treinamento.

Nesse segundo momento, procuramos escolher uma fonte de informação diferente. Como nosso foco são textos que buscam informar a cada gênero especificamente, realizamos uma busca no *Google*[11] com palavras-chave referentes a nomes de grandes sites em português seguido da palavra chave “mulher” e o mesmo foi feito com a palavra-chave “homem”. Ao buscar por “terra mulher” e terra “terra homem”, encontramos na pesquisa os resultados das figuras a seguir.



**Figura 2.1:** Busca no google pelos termos “terra mulher”



**Figura 2.2:** Busca no google pelos termos “terra homem”

O que nos chamou atenção, pois quando inserimos gênero na busca somos imediatamente remetidos a seção de viad e estilo do portal. Além disso, as palavras escolhidas para explicar o conteúdo de cada site são termos popularmente associados ao estereótipo do gênero feminino (amor e sexo, comportamento, casamento, passo-a-paso de beleza) e ao estereótipo do gênero masculino (carros, esportes, lindas gatas, finanças).

Decidimos entrão extrair os textos de colunas da seção de vida e estilo do site [31] para a criação do nosso corpora. Coletamos 101 textos da seção vida e estilo, 133 da subseção voltada para mulheres e 135 textos na subseção de vida e estilo voltada para homens. Os textos são colunas bem diversas e no geral sobre saude, sexo, esporte, transexualidade, estética, automóveis e homossexualidade por exemplo. Optamos por colunas pois compreendem a característica de conter a explicitamente a opinião ou ponto de vista do autor, ou seja, contém os juízos de valor de quem a escreve.

### 2.1.7 Ferramentas

Para a captura, processamento e análise dos textos mencionados utilizamos as seguintes ferramentas:

#### Python

*Python* [9] é uma linguagem de programação de alto nível criada com o propósito de aumentar a velocidade com que o programador desenvolve o código e aumento da produtividade de tal. É uma linguagem de fácil compreensão, aspecto que era primordial no trabalho, uma vez que foi necessário utilizar frameworks para todo o processamento do corpora. Python tem sido muito utilizada em dados científicos e processamento de textos. [39]

#### Scrapy

*Scrapy* [28] é um *framework open source* para a extração de dados de páginas de websites. É desenvolvido em *Python*, e assim como a linguagem, é um *framework* de fácil utilização e rapidez no *web-crawling*.

#### Natural Language Toolkit - NLTK

*NLTK*[22] é um conjunto de ferramentas voltada para o processamento de texto por máquinas que alia tópicos em linguística computacional, tornando o PNL muito mais prático. Desenvolvida também em *Python*, a ferramenta vem sendo chamada de “uma maravilhosa ferramenta para ensino, e trabalho, em linguística computacional” e “uma incrível biblioteca para brincar com linguagem natural”.

## 2.2 Desenvolvimento Prático

Nessa seção iremos descrever o processo de extração parseamento e análise dos dados coletados.

### Extração

Para capturar os textos da web, criamos um *web crawler* a partir do [28], uma aplicação que por meio de um bot extrai os dados do html da página requisitada. Criamos duas classes *Spider*:

- **LinksTerraSpider** a classe captura as urls das página principais que escolhemos.
- **TerraTextSpider** a classe salva os textos das colunas em diretórios baseados nas páginas principais das quais as urls foram extraídas na classe anterior.

```
1 class LinksTerraSpider(scrapy.Spider):
2
3     name = 'linksterra'
4     global_dpage = 1
5
6     def start_requests(self):
7         urls = ['https://www.terra.com.br/vida-e-estilo/', 'https://www.
8             terra.com.br/vida-e-estilo/homem', 'https://www.terra.com.br/
9             vida-e-estilo/mulher']
10
11        for url in urls:
12            yield scrapy.Request(url=url, callback=self.parse)
13
14        def parse(self, response):
15
16            if re.findall(woman, response.url):
17                name_subject = 'terra_vida_estilo_mulher.txt'
18                save_path = 'extracted_texts/'
19                file_name = os.path.join(save_path, name_subject)
20
21                with open(file_name, 'a') as f:
22                    for url in \
23                        response.xpath('//a[@class="main-url text"]/@href')
24                            f.write(str(url))
25                            ).extract():
26                        f.write('\n')
27
28            if next_page is not None:
29                next_page = response.urljoin(next_page)
30                yield scrapy.Request(next_page, callback=self.parse)
```

## Exemplo de texto extraído:

1 título: Já para a piscina: 4 razões para você começar a nadar

2 Geralmente relacionamos a a mergulhos refrescantes num dia quente de verão. Mas basta falarmos de que a galera torce o bico. Mas nós estamos aqui para passar uma mensagem importante: não se deixe levar por essas ideias ruins sobre! ‘‘A natação é um ótimo cardio, completamente sem impacto, e queima calorias do corpo todo’’, afirma Earl Walton, técnico de natação e fundador/diretor do Tailwind Endurance, em Nova York, EUA. Pensando nisso, ele apontou alguns bons benefícios da atividade, que talvez ajudem a te convencer a aderir a modalidade. Você já deve ter ouvido isso antes: para otimizar os treinos, é essencial misturar um pouco as coisas. E com a natação sendo tão diferente dos exercícios na academia, no estúdio ou ao ar livre, ela é ideal para trabalhar os músculos que têm sido negligenciados sem querer. Que nadar pode fortalecer seus braços, pernas, ombros e glúteos, é óbvio. Mas poucos sabem do seu poder sobre o core. ‘‘A água é 724 vezes mais densa do que o ar e providencia uma resistência constante e consistente que força o corpo inteiro a se envolver para conseguir algum impulso’’, explica Walton. Com isso, o core, que segura tudo junto, ganha um tônus incrível no processo.

3 Nadar traz um benefício enorme para o seu condicionamento aeróbico também. O exercício força seu corpo a aprender um novo padrão de respiração. ‘‘Ao expandir a profundidade e o volume de cada respiração, você força os pulmões e o coração a processar o oxigênio com mais eficiência’’, afirma Walton. Dessa forma, o músculo do seu coração ficará mais forte. A natação impulsiona bastante a confiança. Aprender ou aprimorar uma nova habilidade, superando os nervos associados aos potenciais riscos e perceber que você está totalmente sozinho na água é extremamente empoderador. ‘‘Estabelecer uma meta de nadar a piscina toda ou mais pode ser uma experiência que mudará sua vida’’, explica Walton. ‘‘É o que eu sempre vejo nesses anos como treinador e, na minha opinião, é possivelmente o maior benefício que a natação pode dar às pessoas’’.

## Parseamento dos textos

Utilizamos o NLTK[22] para o processamento dos textos extraídos. Como obtemos três classes: *neutra*, criada a partir de textos da página principal, que não é voltada para nenhum gênero, *mulher* e *homem* baseado nos textos encontrados nas páginas voltadas para cada um dos sexos. Essa formatação dos dados foi feito no analisador de viés no código fonte, *bias\_analysis*. Para cada texto lido pelo nosso script em *python* utilizando o *nltk*, seja para o treino, o teste ou a classificação foi feito o parseamento de suas sentenças, transformamos o restante das palavras em tokens e removemos as *stopwords*, pois queremos fazer uma análise que se baseia na probabilidade da palavra estar presente em cada uma das três classes mencionadas.

## Classificação

Buscamos por vieses na linguagem, é possível que palavras que podem ser associadas a estereótipos de gênero sejam mais comumente encontradas em textos que se referem a tal gênero. Assumindo que essa relação entre palavras e estereótipos pode ser detectada, optamos por um algoritmo que classifica dados em grupos aos quais esses dados são mais prováveis de pertencer. Escolhemos o classificador Naive Bayes com aprendizado supervisionado, uma vez que os corpora encontrados na literatura sobre PNL, que poderiam ser utilizados para criar nossas bases de treino, são gerado em contextos específicos e podem não corresponder

à significação das mesmas palavras no nosso corpora. Utilizamos o Naive Bayes para analisar os textos uma vez que supõe que a probabilidade de toda palavra acontecer no texto é independente de outra palavra aparecer, porém a distribuição de palavras como um todo depende da classe a qual o texto pertence. Essa distribuição é calculada durante o treinamento do classificador a partir do corpora.

Como base de treino utilizamos a própria classificação do site Terra, anteriormente mencionada para dividir as classes neutra, mulher e homem. Os tokens obtidos no parseamento do texto foram convertidos em dicionários cuja as chaves são as palavras do texto e os valores associados a elas são *True*. Esse dicionário é chamado de *feature set* e o input para o treinamento do classificador é uma lista de tuplas *feature set* e *label* chamada de *labeled feature sets*. Para cada texto é criado um *feature set* com seu correspondente label. Uma parte dos textos é utilizada no treinamento e a outra é passada para o classificador sem *label* é averiguado se houve o acerto referente a qual classe pertence o texto.

O processo de treinamento e teste foi feito várias vezes, embaralhando a lista dos textos de entrada, o que muda a base de treino a cada vez. Pro resultado final utilizamos uma matriz de confusão com os resultados obtidos somados de várias dessas iterações.

## Resultados

Analisando o resultado do classificador a partir da matriz de confusão verificamos que a tendência é classificar na categoria mulher. Vemos também que os textos que ele classificou como homem ou unissex foram classificados corretamente numa proporção muito maior que os classificados como mulher. Em todos os casos as classificações corretas são maioria. No entanto o viés em classificar o texto como mulher faz com que as tentativas de classificação de homem e unissex tenham tido um baixo índice de acerto. Quase todos os textos relativos a mulher foram classificados corretamente.

1			u
2		m	n
3	h	u	i
4	o	l	s
5	m	h	s
6	e	e	e
7	m	r	x
8			
9	homem	<1491> 3359	150
10	mulher	203 <4616>	181
11	unissex	14 839	<1147>
12			
13	(row = reference ; col = test)		

A partir da tabela de características mais importantes do classificador vemos que a quantidade de palavras que aparecem majoritariamente numa determinada classe não é igualmente distribuída entre as classes. Dentre as 90 mais importantes características apenas 18 são palavras majoritariamente encontrar em textos da classe homem, para mulher 41 são e para unissex são 35. Por outro lado a classe homem possui mais palavras fracamente associadas a essa classe. Entre algumas palavras que chamam atenção como *contudo* está relacionada a mulher quando porém é uma conjunção adversativa com função estrutural e não deveria se relacionar a uma única classe. É possível que uma base de textos maior essa proporção de aparições dessa palavra em textos da classe mulher diminua. Em outro exemplo, palavras como *médica*, *médicos* e *sintomas*, ligadas à classe mulher, estão entre as maiores correlações apresentadas na tabela aparecendo uma razão de 6.3 vezes mais em textos da classe mulher a homem. Isso indica que pode existir uma correlação entre a área da

saúde e a classe mulher. A mesma coisa acontece com a área de alimentação, pois observando as palavras da lista vemos as que são associadas a tal área são mais correlacionadas com mulher.

Esses resultados mostram que é possível identificar relações entre temas específicos e gênero e possivelmente essa abordagem pode obter resultados mais expressivos aumentando-se a base de dados a partir de mais fontes de comunicação.

Rank	Feature	Gender	Score	Baseline
1	Most Informative Features			
2	homens = True	homem : unisse =	11.4	1.0
3	dez = True	unisse : mulher =	8.3	1.0
4	energia = True	mulher : homem =	8.3	1.0
5	2018 = True	unisse : homem =	7.7	1.0
6	assunto = True	homem : unisse =	7.0	1.0
7	local = True	unisse : homem =	7.0	1.0
8	Eles = True	unisse : homem =	7.0	1.0
9	falam = True	homem : unisse =	7.0	1.0
10	respeito = True	mulher : unisse =	7.0	1.0
11	medida = True	mulher : unisse =	6.3	1.0
12	médica = True	mulher : homem =	6.3	1.0
13	médicos = True	mulher : homem =	6.3	1.0
14	comer = True	mulher : homem =	6.3	1.0
15	jornal = True	homem : unisse =	6.3	1.0
16	roupas = True	mulher : unisse =	6.3	1.0
17	sintomas = True	mulher : homem =	6.3	1.0
18	Natal = True	unisse : homem =	6.3	1.0
19	2019 = True	unisse : homem =	6.3	1.0
20	Contudo = True	mulher : unisse =	6.2	1.0
21	publicação = True	unisse : homem =	5.8	1.0
22	sentir = True	mulher : homem =	5.8	1.0
23	motivo = True	homem : unisse =	5.8	1.0
24	nesta = True	unisse : mulher =	5.8	1.0
25	sinais = True	mulher : unisse =	5.7	1.0
26	principais = True	homem : mulher =	5.7	1.0
27	demais = True	mulher : unisse =	5.7	1.0
28	torna = True	unisse : homem =	5.7	1.0
29	compartilhada = True	unisse : mulher =	5.7	1.0
30	Daily = True	homem : unisse =	5.7	1.0
31	leite = True	unisse : homem =	5.7	1.0
32	cabeça = True	homem : unisse =	5.7	1.0
33	tentando = True	unisse : homem =	5.7	1.0
34	Janeiro = True	unisse : mulher =	5.7	1.0
35	entrevistadas = True	homem : mulher =	5.7	1.0
36	estudos = True	mulher : unisse =	5.7	1.0
37	havia = True	unisse : homem =	5.7	1.0
38	Essa = True	unisse : homem =	5.7	1.0
39	melhores = True	homem : unisse =	5.4	1.0
40	EUA = True	mulher : homem =	5.4	1.0
41	água = True	mulher : homem =	5.4	1.0
42	tirar = True	mulher : homem =	5.4	1.0
43	cachorros = True	unisse : mulher =	5.0	1.0
44	britânico = True	homem : mulher =	5.0	1.0
45	pele = True	mulher : unisse =	5.0	1.0
46	campanha = True	unisse : homem =	5.0	1.0
47	animais = True	unisse : mulher =	5.0	1.0
48	população = True	unisse : homem =	5.0	1.0
49	talvez = True	homem : unisse =	5.0	1.0
50	quarta-feira = True	unisse : homem =	5.0	1.0
51	festa = True	unisse : homem =	5.0	1.0
52	gordura = True	mulher : unisse =	5.0	1.0

53	longe = True	mulher : unisse =	5.0 : 1.0
54	receber = True	mulher : homem =	5.0 : 1.0
55	publicou = True	unisse : mulher =	5.0 : 1.0
56	dizem = True	mulher : unisse =	5.0 : 1.0
57	cedo = True	mulher : unisse =	5.0 : 1.0
58	razões = True	mulher : homem =	5.0 : 1.0
59	alimentação = True	mulher : homem =	5.0 : 1.0
60	capacidade = True	mulher : unisse =	5.0 : 1.0
61	disponíveis = True	unisse : mulher =	5.0 : 1.0
62	seis = True	unisse : homem =	5.0 : 1.0
63	próxima = True	mulher : homem =	5.0 : 1.0
64	leve = True	unisse : homem =	5.0 : 1.0
65	massa = True	unisse : mulher =	5.0 : 1.0
66	pequenos = True	unisse : homem =	5.0 : 1.0
67	modelos = True	homem : mulher =	5.0 : 1.0
68	País = True	unisse : mulher =	5.0 : 1.0
69	animal = True	unisse : mulher =	5.0 : 1.0
70	atualmente = True	unisse : mulher =	5.0 : 1.0
71	considera = True	unisse : homem =	5.0 : 1.0
72	40 = True	unisse : mulher =	5.0 : 1.0
73	estresse = True	mulher : homem =	5.0 : 1.0
74	estômago = True	mulher : unisse =	5.0 : 1.0
75	sociedade = True	mulher : unisse =	5.0 : 1.0
76	máximo = True	mulher : homem =	5.0 : 1.0
77	deseja = True	homem : unisse =	5.0 : 1.0
78	cama = True	mulher : unisse =	5.0 : 1.0
79	médico = True	mulher : homem =	5.0 : 1.0
80	nada = True	mulher : unisse =	5.0 : 1.0
81	foto = True	mulher : homem =	5.0 : 1.0
82	mão = True	mulher : unisse =	5.0 : 1.0
83	sensação = True	unisse : mulher =	5.0 : 1.0
84	minutos = True	unisse : homem =	4.7 : 1.0
85	risco = True	mulher : homem =	4.6 : 1.0
86	porque = True	mulher : homem =	4.5 : 1.0
87	teste = True	mulher : homem =	4.3 : 1.0
88	sala = True	unisse : homem =	4.3 : 1.0
89	fique = True	mulher : homem =	4.3 : 1.0
90	rosto = True	homem : mulher =	4.3 : 1.0
91	parece = True	homem : unisse =	4.3 : 1.0

# Capítulo 3

## Conclusões

A diminuição no número de mulheres nas áreas majoritariamente ocupadas por homens, tem como uma das causas a construção social, que abrange cultura, linguagem e estereótipos relacionados ao gênero, aspectos que o indivíduo é exposto desde o nascimento. Em contrapartida, a Inovação de Gênero pode revolucionar a maneira como a ciência é feita, uma vez que as instituições se comprometam a fomentá-la. É um esforço constante para aumentar a diversidade na área, levantando questionamentos e sugerindo novas abordagens, como a Correção de Conhecimento.

Corrigir a maneira como a informação é passada implica em corrigir também os vieses na linguagem, que é parte formadora do ser social e portanto não deve ser enviesada, caso contrário, a perpetuação do estigma da mulher e os estereótipos ligados ao gênero feminino continuarão a ser transmitidos. Nossa análise mostrou alguns traços na linguagem que parecem relacionar termos a um gênero específico. Isso nos motiva a continuar pesquisando e aumentar o nosso corpus, a fim de adicionar mais dados à análise e obter resultados mais expressivos.

Apesar de todos os entraves enfrentados pela mulher na ciência e sociedade, elas seguem firmes lutando pelo seu espaço. Terminamos este trabalho com esperança de que no futuro possa finalmente existir equidade de gênero em todos os meios sociais.



# Referências Bibliográficas

- Ayelet Baram-Tsabari(2010)** Anat Yarden Ayelet Baram-Tsabari. Quantifying the gender gap in science interests. *International Journal of Science and Mathematics Education*. Citado na pág. 7
- Bueno(2015)** Ana Lúcia Dacome Bueno. A produção do sexismo na linguagem: Gênero e poder em dicionários da língua portuguesa. Citado na pág. 10
- Butler(2003)** Judith Butler. *Problemas de gênero: Feminismo e subversão da identidade*. Civilização Brasileira, 00 edição. Citado na pág. 4
- Comissão Europeia(2012)** Comissão Europeia. Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation. Relatório técnico, Comissão Europeia. Citado na pág. 1
- Costa(2015)** Camilla Costa. Estereótipo de que “matemática é para garotos” afasta meninas da tecnologia, diz pesquisador, 2015. URL [https://www.bbc.com/portuguese/noticias/2015/09/150831\\_entrevista\\_andrew\\_meltzoff\\_cc](https://www.bbc.com/portuguese/noticias/2015/09/150831_entrevista_andrew_meltzoff_cc). Citado na pág. 5, 7
- Cunha(2012)** Adan Phelipe Cunha. A emergência da hipótese do relativismo linguístico em Edward Sapir (1884-1939). Dissertação de Mestrado, Faculdade de Filosofia, Letras e Ciências Humanas (Universidade de São Paulo). Citado na pág. 9
- da Silva e Ribeiro(2014)** Fabiane Ferreira da Silva e Paula Regina Costa Ribeiro. A participação das mulheres na ciência: Problematizações sobre as diferenças de gênero. *Ciênc. educ. (Bauru)*. Citado na pág. 6
- de Santana(2012)** Joelton Duarte de Santana. Língua, cultura e identidade: a língua portuguesa como espaço simbólico de identificação no documentário: Língua vidas em português. *Linha D'Água*. URL <http://www.revistas.usp.br/linhadagua/article/view/37367>. Citado na pág. 9
- Foundation(2019)** Python Software Foundation. Python, 2019. URL <https://www.python.org/>. Citado na pág. 13
- Goffman(1975)** Erving Goffman. *Estigma Notas sobre a manipulação da identidade deteriorada*. Zahar. Citado na pág. 4
- Google(2019)** Google. Google, 2019. URL <https://www.google.com/>. Citado na pág. 12
- Harding(1998)** Sandra Harding. Women, science, and society, 1998. URL <http://science.sciencemag.org/content/281/5383/1599>. Citado na pág. 7
- Helmer M e D(2017)** Neef A Helmer M, Schottdorf M e Battaglia D. Gender bias in scholarly peer review. *eLife*. Citado na pág. 6

- J(2017)** Berg J. Looking inward at gender issues. *Science*. Citado na pág. 6
- J. Mattoso Câmara Jr(1955)** J. Mattoso Câmara Jr. Língua e cultura. *Revista Letras - ISSN 0100-0888 (versão impressa) e 2236-0999 (versão eletrônica)*, páginas 1–9. Citado na pág. 9
- Larrosa(2002)** Jorge Larrosa. Notas sobre a experiência e o saber da experiência. *Revista Brasileira de Educação*, página 21. URL <http://www.scielo.br/pdf/rbedu/n19/n19a02.pdf>. Citado na pág. 10
- Marilia Sá Carvalho(2015)** Claudia Medina Coeli e Luciana Dias de Lima Marilia Sá Carvalho. Mulheres no mundo da ciência e da publicação científica. *Cad. Saúde Pública vol.34 no.3 Rio de Janeiro*. URL [http://www.scielo.br/scielo.php?pid=S0102-311X2018000300101&script=sci\\_arttext&tlng=pt](http://www.scielo.br/scielo.php?pid=S0102-311X2018000300101&script=sci_arttext&tlng=pt). Citado na pág. 6
- Narloch()** Leandro Narloch. Mito: “as mulheres são menos produtivas que os homens”. URL <https://veja.abril.com.br/blog/cacador-de-mitos/mito-8220-as-mulheres-sao-menos-produtivas-que-os-homens-8221/>. Citado na pág. 5
- Organização das Nações Unidas(2015)** Organização das Nações Unidas. Transforming our world: the 2030 agenda for sustainable development. Relatório técnico. Citado na pág. 1
- Phillips(2014)** Katherine W. Phillips. How diversity makes us smarter. *Scientific America*. URL <https://www.scientificamerican.com/article/how-diversity-makes-us-smarter/>. Citado na pág. 8
- Presse(2018)** France Presse. Faculdade de medicina no japão baixa notas de mulheres para limitar ingresso delas na instituição, 2018. URL <https://www1.folha.uol.com.br/ciencia/2018/08/faculdade-do-japao-fraudou-notas-para-admitir-menos-mulheres-em-medicina.shtml>. Citado na pág. 5
- Project(2019)** NLTK Project. Nltk natural language toolkit, 2019. URL <http://www.nltk.org/>. Citado na pág. 13, 15
- Santos(2018)** Carolina Marins Santos. Por que as mulheres desapareceram dos cursos de computação?, 2018. URL <https://jornal.usp.br/universidade/por-que-as-mulheres-desapareceram-dos-cursos-de-computacao>. Citado na pág. 1
- Sapir(1921)** Edward Sapir. *Language: An Introduction to the Study of Speech*. Harcourt Brace. Citado na pág. 9
- Sayão(2003)** Déborah Thomé Sayão. Pequenos homens, pequenas mulheres? meninos, meninas? algumas questões para pensar as relações entre gênero e infância. página 78. Citado na pág. 5
- Schiebinger(2008)** Londa Schiebinger. *Gendered Innovations in Science and Engineering*. Stanford University Press. Citado na pág. 7
- Schoellkopf(2014)** Karen Schoellkopf. Hire more women in tech a few simple things you can do to find and hire more qualified women in tech, 2014. URL <https://www.hiremorewomenintech.com/>. Citado na pág. 8

- Scrapinghub e contributors(2019)** Scrapinghub e contributors. Scrapy an open source and collaborative framework for extracting the data you need from websites, 2019. URL <https://scrapy.org/>. Citado na pág. 13, 14
- Shen(2013)** Helen Shen. Mind the gender gap. *nature*. Citado na pág. 7
- Stanford(2019)** Stanford. Gendered innovatios, 2019. URL <https://genderedinnovations.stanford.edu/>. Citado na pág. 7
- Terra(2019)** Terra. Portal terra, 2019. URL <https://www.terra.com.br/>. Citado na pág. 13
- Textio(2019)** Textio. Textio, 2019. URL <https://textio.com/>. Citado na pág. 5
- through Gender(2013)** Innovation through Gender. Gendered innovations how gender analysis contributes to research. Relatório técnico, European Commision. Citado na pág. 7, 8
- Topaz CM(2016)** Sen S Topaz CM. Gender representation on journal editorial boards in the mathematical sciences. *PLoS One*. Citado na pág. 6
- UK Research and Innovation(a)** UK Research and Innovation. Marie curie, a. URL <https://www.ukri.org/women-in-research-and-innovation/gallery-of-pioneers/marie-curie/>. Citado na pág. 3
- UK Research and Innovation(b)** UK Research and Innovation. Grace hopper, b. URL <https://www.ukri.org/women-in-research-and-innovation/gallery-of-pioneers/grace-hopper/>. Citado na pág. 3
- UK Research and Innovation(c)** UK Research and Innovation. Ada lovelace, c. URL <https://www.ukri.org/women-in-research-and-innovation/gallery-of-pioneers/ada-lovelace/>. Citado na pág. 3
- Wikipedia(2019a)** Wikipedia. Microaggression, 2019a. URL <https://en.wikipedia.org/wiki/Microaggression>. Citado na pág. 6
- Wikipedia(2019b)** Wikipedia. Python, 2019b. URL <https://pt.wikipedia.org/wiki/Python>. Citado na pág. 13
- Woods(2017)** Jilian Woods. What statistics show about women in science, 2017. URL <http://blogs.lse.ac.uk/gender/2017/08/21/what-statistics-show-about-women-in-science/>. Citado na pág. 7