



UNIVERSIDADE DE BRASÍLIA - UNB
INSTITUTO DE LETRAS - IL
DEPARTAMENTO DE LÍNGUAS ESTRANGEIRAS E TRADUÇÃO - LET
CURSO DE LÍNGUAS ESTRANGEIRAS APLICADAS – LEA - MSI

JEFERSON VIEGAS RODRIGUES

**UM OLHAR SOBRE GÊNEROS FEMININO E MASCULINO EM RELAÇÕES
TEXTO-IMAGEM RESULTADOS DE TRADUÇÕES AUTOMÁTICAS DE DOCU-
MENTOS MULTIMODAIS – UM ESTUDO EXPLORATÓRIO EM JORNAIS DE
NOTÍCIAS DE LÍNGUA INGLESA**

BRASÍLIA - DF

2019

JEFERSON VIEGAS RODRIGUES

**UM ESTUDO DE GÊNERO EM RELAÇÕES TEXTO-IMAGEM DE RESULTADOS
DE TRADUÇÕES AUTOMÁTICAS DE DOCUMENTOS MULTIMODAIS – UM ES-
TUDO EXPLORATÓRIO EM JORNAIS DE NOTÍCIAS DE LÍNGUA INGLESA**

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI), sob orientação do Prof. Dr. Thiago Blanch Pires, da Universidade de Brasília (UnB).

BRASÍLIA - DF

2019

JEFERSON VIEGAS RODRIGUES

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Línguas Estrangeiras Aplicadas ao Multilinguismo e à Sociedade da Informação (LEA-MSI), sob orientação do Prof. Dr. Thiago Blanch Pires, da Universidade de Brasília (UnB).

Aprovado em ___/___/___

Prof. Dr. Thiago Blanch Pires
Universidade de Brasília
Orientador

Prof^a. Dra. Elisa Duarte Teixeira (LET/UnB)
Universidade de Brasília
Avaliadora

Prof^a. Dra. Fernanda Alencar Pereira
Universidade de Brasília
Avaliadora

AGRADECIMENTOS

Inicialmente eu chamo um agradecimento ao universo, ele com suas vibrações e sintonia me permite ver caminho e desbravar estradas que nem mesmo imaginava que poderiam existir, simplesmente por estar bem consigo mesmo e sempre dizer obrigado pelas suas conquistas grandes ou pequenas.

Agradeço muito meus pais, Rose e Lezimar, que apesar de ficarem surpresos com minha decisão da segunda graduação, me apoiaram e deram forças pra seguir, e de me ensinarem a ser esse quem eu sou, persistente e obstinado. Agradeço bastante também ao meu irmão David por ter sido a pessoa a plantar a sementinha do LEA em mim e super me apoiado nesse percurso, desde o dia da matrícula até nos momentos mais desesperados do TCC.

O meu obrigado ao Professor Thiago que mais que orientador, transcendeu esse papel olhando o lado humano e me apoiando em momentos de desespero, além de ter mostrado pra mim caminhos a seguir. A Professora Fernanda, não só por ter aceitado o convite para fazer parte da banca, mas também pelos conselhos, conversas e ajudas ao longo do curso. À Professora Elisa que além de uma professora incrível, é uma pessoa incrível e eu me sinto muito honrado por tê-la como parte da banca, assim como também não poderia deixar de agradecer ao COMPLETT, que sob sua chefia e a orientação do professor Thiago me proporcionou várias oportunidades esse ano, como a indicação a um prêmio destaque inesperado e uma viagem a um congresso maravilhoso em um dos lugares mais bonitos do Brasil, João Pessoa.

UnB, LEA! Obrigado por terem me proporcionado a oportunidade de um curso que eu realmente me encontrei e que me possibilitou ver que o mundo ainda é muito grande e precisa ser desbravado.

Agradecimento especial a Professora Helena, que eu considero minha mestra, tutora na legendagem algo que, além de ser prazeroso de se fazer é uma das minhas qualificações e eu fico muito honrado de ter aprendido com ela. Além dela, ao Professor Marcos Carneiro, por toda parceria, oportunidades e ajuda que me proporcionou.

À Érica que está comigo desde que éramos moleques, embarcou comigo na doideira de um segundo curso e até hoje está na minha vida, em momentos bons e ruins, muito obrigado por todo apoio e força ao longo de todos esses anos.

À todas as amigas maravilhosas que fiz ao longo dessa graduação: Anna, obrigado pelos tantos conselhos, risadas, choros, ajudas, isso não teria sido nem metade do que foi sem você. Augusto, obrigado por todos os momentos que estive comigo, por tanto que me suportou

nos meus momentos mais chatos de surto, por todos os Pokémon! Anahy, com seu riso fácil que contagia a gente, pelos papos, pela ajuda no espanhol. Cris, valeu pela companhia desde o primeiro dia de aula e por todo carinho. Renan e Janaína, que mesmo tendo gostos e personalidades tão diferentes dos meus, mostram que uma amizade quebra essas barreiras. Sofia, lá em Minas e a gente sentindo falta aqui, obrigado por todos os momentos de diversão, ajudas e pelas caras indignadas por uns certos Pokémon brilhantes. Lucas, uma pessoa que só fui tirar uma dúvida e se tornou alguém por quem tenho um carinho muito especial hoje, obrigado por me ouvir, pelos conselhos e me fazer rir em momentos sombrios. Max, por uma amizade que começou com conversas sobre escrita e culminou no que é hoje, obrigado a você por toda força em momentos tristes, toda ajuda, pelo riso frouxo que é mais engraçado que a piada às vezes, agradeço pelos momentos. Geovana, que nem sei muito bem como foi, mas foi e hoje é alguém por quem tenho um carinho muito grande. Assim como a Victória, que mesmo querendo me bater as vezes, eu gosto demais, obrigado por me incentivar no estudo de Libras e acessibilidade. Letícia, valeu pelos papos indignados e revoltados, compreensão e ajuda que me deu. Rafael e Marion que mesmo do outro lado do Atlântico, estão presentes em minha vida, sempre me apoiando e me ajudando, seja me incentivando a escrever ou me dando forças pra continuar. Aline e Gustavo, que me ajudaram demais, demais, demais no período de Gestão do CA, onde cresci muito como pessoa ao lado deles, obrigado por tudo. Thales, que ajudou com diversão em momentos que eu mais precisava e acabou se tornando alguém especial pra mim, obrigado! A pessoinhas que mesmo na reta final, não achei que fosse ter uma amizade assim, mas obrigado também, Cléber e Louise.

À Ana Laura, minha maravilhosa psicóloga que ajudou a suportar momentos difíceis que tenho certeza que não conseguiria por conta própria, além de ter sido um anjo em minha vida em uma das épocas mais difíceis.

Espero que não tenha ficado ninguém de fora! Mais uma vez agradeço a todas e a todos vocês que me ajudaram a crescer e nessa trajetória tão louca e cheia de altos e baixos que foram esses quase cinco anos do LEA na UnB.

RESUMO

Há muitas pesquisas e estudos em desenvolvimento sobre erros de tradução automática. Porém, um campo pouco explorado é o da tradução automática envolvendo sexismo na língua e que levem em consideração esses erros em documentos multimodais. Nesse sentido, o objetivo do presente trabalho foi explorar esses documentos com figuras de mulheres em destaque e levantar a questão do sexismo presente nos erros de tradução. Baseando-se no trabalho de Rodrigues e Pires (2019), onde a problemática deste estudo foi inicialmente levantada. Documentos formados por texto-imagem foram utilizados para elaboração da tradução automática com o Google Tradutor, para obtenção do banco de dados analisado. Resultados mostraram uma grande tendência da tradução automática, em casos de ambiguidade de gênero, optando quase sempre por frases e palavras no masculino para passar o texto para o português. Objetivou-se levantar esse problema pouco explorado no campo da tradução automática e percebeu-se que o sexismo está na língua e no tradutor automático, pois está presente na sociedade, de modo que essas ferramentas têm a tendência de absorver tais comportamentos sociais.

Palavras-chave: Erros de Tradução Automática. Google Tradutor. Multimodalidade. Gênero. Sexismo.

ABSTRACT

There is a lot of research and studies under development on machine translation errors. However, a field which has not almost any exploration is that of machine translation involving sexism in language which also take into account these errors in multimodal documents. In this sense, the objective of the present work was to explore documents with prominent women figures and to analyze the issue of sexism present in translation errors. It arose based on the study by Rodrigues and Pires (2019), where the problem of this study was initially approached. Text-image Documents were machine translated with Google Translate to obtain the analyzed database. Results showed a great tendency of machine translation, in cases of gender ambiguity, to choose phrases and words in masculine to render the text into Portuguese. The objective was to show this little explored problem in the field of machine translation. We noticed that sexism is present in the language and in the automatic translator, because it is present in society, and these tools tend to mimic such social behaviors.

Keywords: Machine Translate Errors. Google Translate. Multimodality. Gender. Sexism.

LISTA DE FIGURAS

Figura 1 - Assange's lawyer/Advogado de Assange	4
Figura 2 - Categorias de textura intersemióticas de Liu & O'Halloran (2009).....	6
Figura 3- Imagem de manifestantes em Hong Kong.....	8
Figura 4 – Categorias de erros propostas por Vilar et al. (2006)	10
Figura 5 - Tradução Automática da Figura 3	11
Figura 6 - Tradução Goalkeeper/Goleiro	13
Figura 7 - Tradução da palavra Nurse para o português	20
Figura 8 - Qandeel Baloch was killed by her brothers	26
Figura 9 - Qandeel Baloch assassinada pelos irmãos	26
Figura 10 - Lyubov Sobol, the invisible influencer.....	28
Figura 11 - Lyubov Sobol, a influenciadora invisível.....	28
Figura 12 - Diversidade cultural com digital influencer	29
Figura 13 - Instagram esconde likes dos digital influencers.....	30
Figura 14 - Ascensão da Digital Influencer Júlia Lawrence	30
Figura 15 - Russian agent, Maria Butina.....	31
Figura 16 - Agente Russa, Maria Butina.....	32
Figura 17 - Samantha Kureya, comedian from Zimbabwe	33
Figura 18 - Samantha Kureya, comedianta do Zimbábue	34
Figura 19 - Sentenças do Húngaro para o Inglês.....	35
Figura 20 - Capturas de tela do Google Tradutor.....	36
Figura 21 - Claus von Bülow, Socialite	37
Figura 22 - Claus von Bülow, Socialite	37
Figura 23 - Ksenia Sobchak, Socialite	39
Figura 24 - Ksenia Sobchak, Socialite	39

LISTA DE TABELAS E QUADROS

Tabela 1 – Polissemia Intersemiótica - Relações Lógicas entre língua e imagens Liu & O'Halloran (2009).....	8
Tabela 2 – Subcategorias de Palavras Incorretas.....	11
Quadro 1 - Séries de exemplificação da relação sexo e gênero na língua.....	16

SUMÁRIO

INTRODUÇÃO	1
1. INCOMPATIBILIDADE INTERSEMIÓTICA	3
1.1 Textura intersemiótica	6
1.1.1 Paralelismo intersemiótico	7
1.1.1.1 Homoespecialidade.....	7
1.1.1.2 Estruturas paralelas intersemióticas.....	7
1.1.2 Polissemia intersemiótica	8
1.1.3 Relações lógicas entre língua e imagens.....	8
2.2 Erros de tradução automática, desvios léxico-semânticos de Vilar et al. (2006)	10
1.2.1 Palavras Faltantes	11
1.2.2 Ordem das palavras.....	12
1.2.3 Palavras incorretas	12
1.2.4 Palavras desconhecidas.....	13
1.2.5 Pontuação	13
2. GÊNERO NO CIBERESPAÇO.....	14
2.1 Sexismo na língua.....	16
2.1.1 Sexo X Gênero	17
3. GOOGLE TRADUTOR.....	19
3.1 Tradução automática e viés de gênero	21
4. METODOLOGIA.....	22
4.1 As Fontes de Dados	22
4.2 Métodos e Ferramentas.....	23
5. ANÁLISE	24
5.1 Qandeel Baloch	25
5.2 Lyubov Sobol	27
5.3 Maria Butina.....	31
5.4 Samantha Kureya “Gonyeti”	33

5.5 Claus von Bülow	36
6. CONSIDERAÇÕES FINAIS.....	41
REFERÊNCIAS BIBLIOGRÁFICAS	43

INTRODUÇÃO

Tradutores automáticos são cada vez mais utilizados no dia-a-dia das pessoas, seja com objetivos pessoais ou profissionais. Assim, vários avanços são enxergados nessa modalidade, porém poucos estudos são feitos para o aperfeiçoamento de tal ferramenta, principalmente no que se refere aos documentos multimodais, ou seja, aqueles que usam mais de um modo - exemplo texto junto de uma imagem - para formar seu significado final.

O início desse estudo se deu no Programa Institucional de Bolsas de Iniciação Científica - PIBIC, da Universidade de Brasília, onde, além dos achados que a pesquisa apresentou, um deles se destacou por ter sido além do esperado. Documentos multimodais com mulheres como figuras de destaque nunca são traduzidos na forma correta, sendo convertidos pelo tradutor automático para a forma masculina em seu documento de chegada.

As razões dessa problemática podem ser várias, mas algumas das hipóteses são as seguintes: o machismo na língua, uma vez que as mulheres no passado eram impedidas de participação em vários setores da sociedade, possivelmente também na formação de normas gramaticais; a cultura machista do mundo refletida em orientações linguísticas do tradutor automático.

Assim, este trabalho propõe um estudo de gênero em relações texto-imagem de resultados de traduções automáticas de documentos multimodais partindo de textos em jornais de notícias em língua inglesa.

O objetivo desta investigação é analisar o gênero masculino e feminino em incompatibilidades intersemióticas geradas em tradução automática de artigos de notícias online. Além disso, analisando quais seriam os impactos que esses erros de tradução em sua forma poderiam causar.

Este estudo tem um caráter exploratório e qualitativo, baseando-se na busca por documentos necessariamente multimodais, em sites de notícias, que apresentem mulheres como figuras de destaque e que, após a tradução automática, apresentem incompatibilidade intersemiótica relacionada ao gênero incorreto nas legendas.

A importância desse estudo está em seu pioneirismo em um campo pouco explorado, que é o da tradução automática de textos intersemióticos, mas ainda no que diz respeito especificamente à questão de gênero. Assim, com o padrão de erros identificado, poderia futuramente ajudar na melhoria de uma das ferramentas de tradução mais utilizadas, o Google Tradutor. E além desse aspecto, exibir o problema do gênero feminino e masculino questão linguística em uma tentativa de desconstrução de preconceitos arraigados na língua.

Na primeira seção, intitulada “Incompatibilidade intersemiótica”, apresentar-se-á uma explanação desse conceito, além da maneira que ele se relaciona com outros como o de textura intersemiótica e a ligação com os desvios léxico-semânticos propostos em estudos.

Em “Gênero no ciberespaço” será levantado e mostrado que, mesmo em um ambiente aparentemente neutro por natureza, ele agrega desigualdades, preconceitos e vieses do mundo real e replica isso no ambiente virtual. Bem como sexismo na língua, que trará mostras de como os movimentos feministas se impuseram para a luta contra questões importantes de sexismo identificadas nas línguas e também uma diferenciação de sexo e gênero e como eles influem na construção de sentenças.

A seguinte trará uma mostra do que é o Google Tradutor e a razão de ele ser objeto de estudo, uma vez que é um tradutor dos mais utilizados e dados sobre seu funcionamento desde a concepção até os dias mais recentes. Em “Tradução automática e viés de gênero”, um histórico da ideia de tradução automática será abordado. Também a dificuldade que esses tradutores possuem com documentos ditos multimodais e os vieses sexistas gerados na tradução automática.

Em seguida, abordar-se-á a metodologia do estudo e a maneira como se coletou os dados, de onde vieram, além de métodos e ferramentas que foram necessários para seu tratamento para a obtenção de resultados. Na sequência a análise desses resultados mostra o sexismo agregado pelo tradutor automático.

Por fim, nas considerações finais deste estudo, o parecer sobre a investigação, às respostas as hipóteses e proposições para o futuro da pesquisa.

1. INCOMPATIBILIDADE INTERSEMIÓTICA

Desde a década de 1980, com mudanças em arquétipos de sistemas anteriormente baseados em regras para sistemas baseados em estatísticas, internacionalmente, graças ao uso do computador e da globalização, tem se expandido a utilização de tradutores automáticos (HUTCHINS, 2010). Não só essa forma de tradução, mas também a utilização de documentos formados por mais do que meramente textos, por exemplo imagens e sons – que tem como objetivo a concepção de um significado construído do resultado da integração de mais de um desses modos (BATEMAN, 2008). Apesar disso, mesmo que os tradutores automáticos sejam aperfeiçoados para realizar a tradução de textos e sons, não se observa um conjunto que abranja as relações intersemióticas resultantes da integração texto-imagem (BATEMAN, 2014).

Nesse contexto de multimodalidade, Liu & O'Halloran (2009) propõem uma abordagem por eles definida como Textura Intersemiótica, “que compreende o uso de diferentes dispositivos coesivos entre texto e imagem para a produção de coerência” (RODRIGUES e PIRES, 2019). Nessa situação, ao se usar um tradutor automático, pode ocorrer a chamada Incompatibilidade Intersemiótica (PIRES, 2017), isto é, uma nova configuração de significado no texto de chegada destoante da proposta originalmente presente naquele texto de partida.

Em Rodrigues e Pires (2019), há um estudo exploratório sobre incompatibilidades intersemióticas geradas por traduções automáticas - inglês > português - realizadas com base no Google Tradutor. Dentre os achados significativos do projeto, um dos que mais se destaca é o fato de, numa construção intersemiótica cuja figura de destaque seja uma mulher e na legenda da imagem do texto de partida em inglês não é especificada que este se refere a uma mulher, o Tradutor sempre optar por uma tradução no masculino no português, como mostra a figura a seguir:

Figura 1- Assange's lawyer/Advogado de Assange



Fonte: Rodrigues e Pires (2019)

A relação intersemiótica é quebrada por apresentar formas incorretas destoantes no texto de chegada em relação ao de partida, uma vez que “advogado” se refere, na verdade, a Jennifer Robinson, uma mulher, então “advogada” seria o termo correto a se adotar.

Essa pesquisa se desenvolverá baseada nesse achado de Rodrigues & Pires (2019), objetivando uma análise mais profunda sobre esse tipo de fenômeno com a utilização da tradução automática. A seguir retrata-se a fundamentação teórica utilizada neste estudo.

Pires (2017), em sua tese intitulada *Ampliando olhares sobre a tradução automática online: um estudo exploratório de categorias de erros de máquina de tradução gerados em documentos multimodais*, já erigiu a questão dos erros de tradução automática (ibid, p. 14):

Desde o século dezessete já se discutia a ideia de se mecanizar o processo tradutório. Mas foi com o advento dos computadores no período do pós-guerra, que novas disciplinas se formaram em torno do desenvolvimento desse conceito (HUTCHINS, 1986, 2010). Assim surgiram, por exemplo, áreas como a Ciência da Informação (CI) e mais especificamente a Linguística Computacional (LC), ambas interessadas na investigação de problemas pontuais de tradução automática, envolvendo ambiguidade lexical, alinhamento e análise semântica.

Ainda que se mostrem relevantes estudos na área e exista interesse em tais perspectiva, é visível que o número de estudos está muito aquém do que se poderia ter. Maior ainda é a escassez quando o documento traduzido é multimodal (PIRES, 2017).

Bateman (2008) declara que esses tipos de documentos podem ser classes documentais podem ser classificadas como uma profusão de “modos” que são visualmente interligados, objetivando formar uma coleção interligada de objetos entrelaçados que se comunicam. Por “modo”, Kress & Van Leeuwen (2006) denominam as formas de comunicação baseadas em recursos de meios diferentes, por exemplo, auditivos e visuais, e os escritos.

Quando esses modos, ao passarem pela tradução automática, geram uma configuração coesiva – ou perda desta – diferente da originalmente proposta no documento de origem, tem-se a chamada “Incompatibilidade Intersemiótica”, que Pires (2017) definiu em sua tese como (p. 107):

A única forma pela qual se pode reconhecer essas “mudanças” de significado visual-verbal na tradução é quando há incompatibilidade (*mismatch*) léxico-semântica referindo-se a uma imagem ou parte dela, ou seja, a imagem não representa os componentes semânticos que faltam na tradução. Por essa razão intitulou-se “incompatibilidade intersemiótica” (*intersemiotic mismatch*) a partir da geração de texto automaticamente produzido.

Como base para levantamento, catalogação e análise de tais incompatibilidades, tomou-se como base os estudos de Liu & O’Halloran (2009) em que trazem a noção de *Textura Intersemiótica*, “sendo esta uma questão de relações semânticas entre diferentes modalidades realizadas através de Dispositivos Coesivos Intersemióticos no discurso multimodal” (RODRIGUES & PIRES, 2019), entendendo-se por esses dispositivos a junção do tema de um documento, sua informação e seu paralelismo, com a coesão de outro para assim formar-se a Textura. Pode-se dizer que é o dispositivo mais importante concernente aos documentos multissemióticos, uma vez que sua função é integrar palavras e imagens, não só uma simples ligação entre dois modos.

Uma importante consideração a se fazer, corroborando Pires (2017), é a necessária diferenciação entre multimodalidade e multissemiótico. Multimodalidade aqui é um conceito amplo, uma área de investigação versus os modos diferenciados de comunicação; multissemiótico, como abordado por Liu & O’Halloran (2009), seria aquele documento que se utiliza de mais de um meio para formação de seu significado.

Outra distinção também necessária é entre Tradução e Relação Intersemiótica. O prefixo “inter”, é utilizado para formação de palavras que se relacionam “entre”, como Internacional – entre nações. Assim, Intersemiótica seria a relação entre as semioses. Porém, quando se fala em

tradução intersemiótica, é em referência a um termo cunhado primeiramente por Roman Jakobson, que trata esse tipo de tradução como uma “transmutação” de um sistema de significação para o outro (PLAZA, 2003); tome-se como exemplo a adaptação de um livro para um filme.

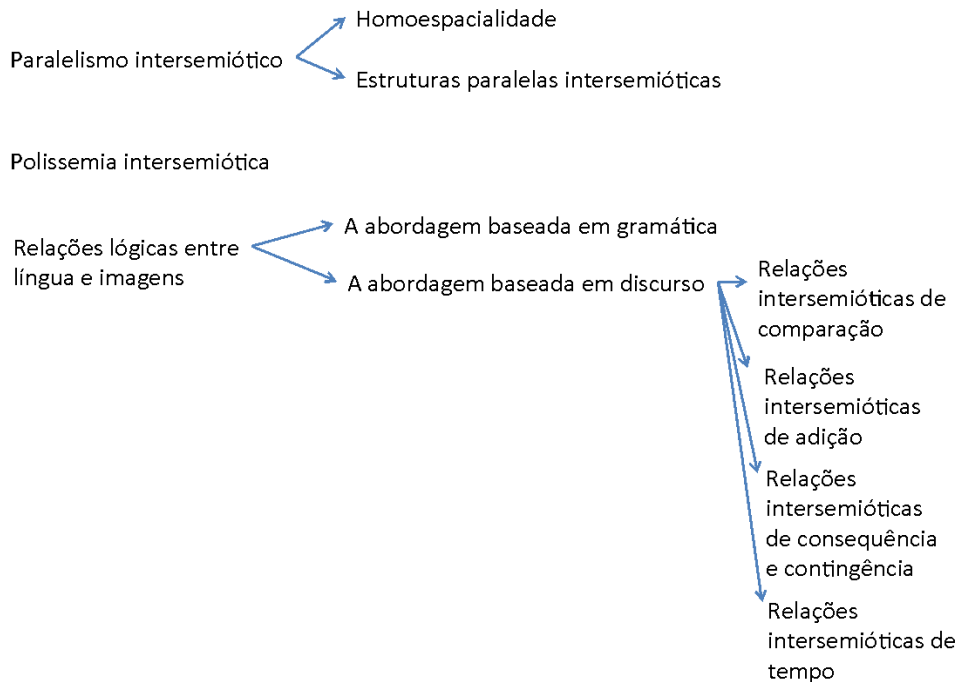
A relação intersemiótica, tratada neste documento como “textura”, salientando os estudos de Liu & O’Halloran (2009), é a maneira como se dá o vínculo entre os modos de um documento multissemiótico. Na seção a seguir, o conceito será aprofundado e mais bem explicado percorrido, antes de se apresentar seu elo com a questão de gênero mostrada na Figura 1.

1.1 Textura intersemiótica

Liu & O’Halloran (2009) definem a textura intersemiótica como uma questão de relações semânticas entre diferentes modalidades realizadas através de Dispositivos Coesivos Intersemióticos no discurso multimodal. É o atributo crucial dos textos multissemióticos que cria a integração de palavras e figuras, em vez de um mero vínculo entre os dois modos.

Dentro desse conceito, há um rol de categorias que pode ser conferido de forma panorâmica na figura a seguir:

Figura 2 - Categorias de textura intersemióticas de Liu & O’Halloran (2009)



1.1.1 Paralelismo intersemiótico

Segundo os Liu & O'Halloran (2009), a primeira dessas grandes categorias, o Paralelismo Intersemiótico, refere-se à relação de coesão responsável por interconectar imagem e texto, por exemplo, onde ambos compartilhariam formas similares. Esta ainda é subdivida em Homoespacialidade, quando a relação toma forma no mesmo plano de expressão, e Estruturas Paralelas, onde os modos apresentam conteúdo semelhante.

1.1.1.1 Homoespecialidade

É uma espécie de paralelismo entre imagem e a língua escrita quando os dois modos estão ocupando o mesmo plano, assim formando o significado da imagem. Em outras palavras, a relação texto-imagem, o texto ocupa o mesmo espaço da imagem, diferenciando-se de uma legenda por exemplo.

1.1.1.2 Estruturas paralelas intersemióticas

Nas estruturas paralelas, texto e imagem vão compartilhar uma semelhança em seu significado para a formação de um sentido final do texto. Na Figura 3, extraída do jornal de notícias *BBC News*, por exemplo, tem-se estruturas paralelas intersemióticas e, ao mesmo tempo, homoespacialidade. A imagem, por si só, dá a entender que está acontecendo um contexto de manifestações, os cartazes fazem parte dela com seus textos ressaltando esse significado e apresentam uma pista de que é uma manifestação em favor da democracia, principalmente pela frase “*power to the people*”¹ em letras garrafais. Fato que é confirmado pela legenda, pois a imagem deixa subentendido o que a legenda confirma paralelamente, descrevendo os fatos que se desdobram: “*Pro-democracy protestors attending a rally in Hong Kong*”².

¹ “Poder para o povo”, tradução do autor.

² “Manifestantes pró-democracia participando de um comício em Hong Kong”, tradução do autor.

Figura 3- Imagem de manifestantes em Hong Kong³



Fonte: BBC News

1.1.2 Polissemia intersemiótica

A “polissemia intersemiótica” (LIU & O’HALLORAN, 2009; p.375) é descrita como uma textura onde a relação de coesão entre os modos verbais e visuais compartilham múltiplos significados em um documento multissemiótico. Em outras palavras, nessa categoria há um compartilhamento de significados similares e diferentes, o que resulta em uma “co-contextualização de relações entre linguagem e imagens e convergência experiencial em textos multissemióticos” (RODRIGUES e PIRES, 2019).

1.1.3 Relações lógicas entre língua e imagens

Em relação a essa terceira e última categoria, Pires (2017, p. 91) elucida que os autores “buscam tratar das relações lógicas entre língua e imagens, isto é, da análise de significados lógicos entre componentes verbais e visuais baseada em dois tipos de abordagem: (i) gramatical e (ii) discursiva”.

A primeira dessas abordagens, Liu & O’Halloran (2009) definem como uma descrição primária de significação lógica entre diferentes recursos semióticos. Sobre abordagem baseada

³ Disponível em: < <https://www.bbc.com/news/newsbeat-49371809> >. Acesso em: 22 de agosto de 2019.

no discurso, também declaram que ela explica as relações de sentenças complexas, isso para que o usuário analisador possa selecionar a unidade analítica intersemiótica apropriada, se libertando das amarras prescritas pela abordagem gramatical.

A seguinte tabela explicita as subdivisões da abordagem baseada no discurso.

Tabela 1 – Polissemia Intersemiótica - Relações Lógicas entre língua e imagens (LIU & O'HALLORAN, 2009)

Relação intersemiótica de comparação	Recurso para organização do significado lógico em relação à semelhança entre linguagem e imagens no discurso multimodal. Assim, quando os componentes visuais e linguísticos compartilham significado experimental semelhante, os diferentes modos são sempre reformulações semióticas entre si e suas relações lógicas serão identificadas como comparativas. Além disso, sempre estarão acompanhadas de dispositivos intersemióticos coesivos.
Relação intersemiótica de Adição	Um componente semiótico, ou modo, adiciona novas informações ao outro e, portanto, as duas mensagens são unidas.
Relações intersemióticas de consequência e contingência	Segundo Pires (2017, p. 94), as relações de consequência referem-se à “causais não-modalizadas entre mensagens verbais e visuais onde o efeito foi assegurado”; já as de contingência estão presentes em textos multissemióticos onde a causa tem potencial para determinar uma possibilidade, mas sem um efeito assegurado.
Relações intersemióticas de tempo	Quando diferentes etapas processuais são representadas verbal e visualmente, as relações imagem-texto podem ser descritas com a lógica do tempo. Geralmente apresentados em manuais de procedimentos, por exemplo, manuais técnicos ou o site de procedimentos <i>WikiHow</i> ⁴ .

Fonte: próprio autor

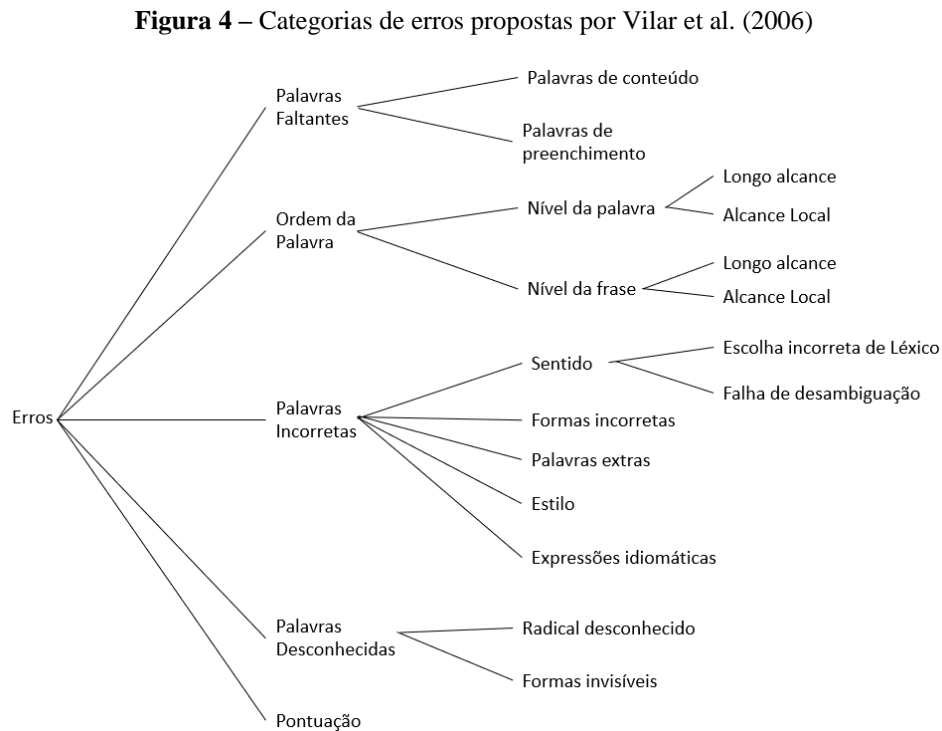
⁴ No link: <<https://pt.wikihow.com/Página-principal>>

2.2 Erros de tradução automática, desvios léxico-semânticos de Vilar et al. (2006)

O trabalho de Vilar et al. (2006) traz um vislumbre sobre a avaliação dos resultados de tradução automática. Como citam os autores, ainda é um trabalho muito controverso, há muitas ferramentas propostas para essas avaliações, porém uma relação entre os erros previstos por elas e os erros reais encontrados nas traduções não é fácil de encontrar. A identificação dos problemas mais importantes de um sistema de tradução é essencial para concentrar os esforços de pesquisa.

Para encontrar os erros em uma tradução, é útil ter uma ou mais traduções de referência para contrastar o resultado da tradução automática com o texto correto. No entanto, como é de conhecimento amplo na comunidade da tradução automática, existem várias traduções consideradas corretas para uma determinada sentença de origem, o que representa um problema difícil na avaliação e comparação de sistemas de tradução automática (VILAR ET AL., 2006).

Assim, com vistas a facilitar a análise desses erros de tradução automática, os autores propõem uma categorização deles como mostra a Figura 4 a seguir:



Fonte: Rodrigues e Pires (2019, p. 7)

1.2.1 Palavras Faltantes

De acordo com Vilar et al. (2006), um erro de palavras faltantes é ocasionado quando alguma palavra não está presente na frase de chegada, em detrimento da frase de partida. A Figura 5, a seguir, tomou a figura 3 como base e passou pelo tratamento de tradução automática do Google Tradutor, e o resultado pode ser conferido a seguir.

Como é possível visualizar, a palavra *protestors*, que em tradução direta para o português seria “manifestantes”, foi totalmente omitida após a TA, o que exemplifica claramente o erro de palavra faltante, além de quebrar a textura de paralelismo intersemiótico contido anteriormente, mesmo ainda tornando a frase em português ausente de nexos, uma vez que ela permanece sem um sujeito definido. Assim, o exemplo ainda poderia se encaixar em uma das subdivisões de palavras faltantes, “palavras de conteúdo”, que são essenciais para manter o significado da sentença, em detrimento da outra divisão, “palavras de preenchimento”, que só estariam ali para manter a sentença gramaticalmente correta, mas mantendo o significado originalmente proposto.

Figura 5 - Tradução Automática da Figura 3



Fonte: BBC News

1.2.2 Ordem das palavras

Nesta categoria (VILAR ET AL., 2006), pode-se distinguir a reordenação das palavras baseada em palavras ou frases, e ambas podendo se subdividir em alcance local e alcance longo. Em se tratando dos rearranjos baseados em palavras, é possível uma movimentação delas de forma individual, sem dependência das outras, à medida que os baseados em frases, devem ser deslocados em blocos de vocábulos consecutivos para não gerar perda do sentido da partida.

1.2.3 Palavras incorretas

Nesta categoria, Vilar et al. (2006) descrevem a classificação mais numerosa de seu trabalho, palavras incorretas. Refere-se a quando não se encontra a palavra apropriada após a TA, o que pode ser subdividido em mais cinco subcategorias.

Tabela 2 – Subcategorias de Palavras Incorretas

Sentido	Ocorre uma quebra do significado da sentença, podendo apresentar mais duas subdivisões: <ul style="list-style-type: none"> • Escolha de tradução incorreta pela TA; • TA incapaz de desambiguar o significado em dado contexto.
Formas incorretas	A tradução não produz uma forma adequada para a palavra, embora de maneira básica a tradução esteja certa.
Palavras extras	Gera vocábulos sobressalentes que desconfiguram a ideia original.
Estilo	Escolha imprópria de palavras na tradução que, embora mantenham o significado, não podem ser consideradas corretas.
Expressões idiomáticas	TA não reconhece uma expressão idiomática e faz a tradução ao pé da letra, o que acarreta erros, já que essas não poderiam ser transpostas dessa maneira a outro idioma.

Fonte: próprio autor

1.2.4 Palavras desconhecidas

Esta categoria é definida por Vilar et al. (2006) de maneira breve, podendo ser dividida em: palavras verdadeiramente desconhecidas – quando o sistema de TA não consegue traduzir para o idioma de chegada o documento proposto. Bem como formas invisíveis, em que a palavra simplesmente não aparece no resultado da tradução (vide Figuras 3 e 5).

1.2.5 Pontuação

Categoria que não é trabalhada por Vilar et al. (2006). Seu levantamento é a título de curiosidade e menção. Isso se dá pois, nas traduções automáticas, esses tipos de erros são considerados de pequeno porte; assim, não são levados em consideração na hora dos estudos de análise de TAs. Desta maneira, assim como fazem os autores em seu estudo, neste o erro também será levado em consideração.

A seguir, na Figura 6, existe uma comparação da legenda que a originalmente a acompanhava com aquela gerada no resultado da TA para o português nos estudos de Rodrigues e Pires (2019), ela mostra um claro erro de desambiguação, além da questão do gênero da tradução, uma vez que, mesmo se tratando da Copa Mundial de Futebol Feminino, a TA do Google Tradutor optou por uma tradução masculina de *goalkeeper*.

Figura 6 - Tradução *Goalkeeper*/Goleiro



Fonte: Rodrigues e Pires (2019, p. 10)

Nos capítulos seguintes se vai discorrer sobre as questões de gênero e tecnologia e uma explanação de algumas concepções delas em algumas ópticas.

2. GÊNERO NO CIBERESPAÇO

Antes de entrar na questão supracitada de gênero relacionado à tradução automática em documentos multissemióticos, faz-se necessária uma contextualização para além dos dispositivos de intersemióticos e de erros de tradução. Isso porque, há caminhos e motivos muito antigos que podem ter acarretado esses desvios de tradução, em se tratando de uma mulher como figura de destaque.

Muito se tem falado de igualdades geradas pelo desenvolvimento das tecnologias da comunicação, graças ao acesso à informação ser mais fácil. No entanto, Nathansohn e Brunet (2009) refutam essa ideia, uma vez que esse desenvolvimento tecnológico horizontal pode ser considerado uma falácia, pois mesmo a internet sendo, em tese, um ambiente propenso à igualdade, graças a democratização e facilidade de acesso, as Tecnologias da Informação e a Comunicação não estão alheias às relações de poder presentes no mundo, o que produz complexas desigualdades entre homens, mulheres, brancos, negros, pobres e ricos.

O que ainda se observa, segundo as mesmas autoras (ibid.; p. 175) é uma exclusão das mulheres desse meio eletrônico de informações. “A exclusão das mulheres de toda proximidade com a computação é histórica, assim como o registro histórico é patriarcal e não reconhece as contribuições das mulheres para as ciências da informática e a matemática”⁵. Essa percepção vem desde muitos anos, e ao final do século XX surgiu um termo através dos movimentos feministas para modificação desse cenário, *HerStory*.

*HerStory*⁶ é um jogo de palavras da língua inglesa com a palavra *History*, história em que *his* (“dele”) é trocado por *her* (“dela”). O pronome possessivo masculino da terceira pessoa do singular, *his*, está presente no começo da palavra, então a ideia que passa é literalmente de “história dele”, ou seja, história apenas contada do ponto de vista masculino. Assim, substituindo pelo pronome possessivo feminino *her*, o que se teria, em uma tradução aproximada,

⁵ Tradução do autor. Original: *La exclusión de las mujeres de toda proximidade a la computación es histórica, así como el registro historiográfico es patriarcal y no reconoce las contribuciones de las mujeres a la ciencia informática y matemática.*

⁶ Definição Urban Dictionary. Disponível em: <<https://www.urbandictionary.com/define.php?term=herstory>>. Acesso em: 09 nov. 2019.

seria “história dela” ou “dona da história”. O mesmo jogo de palavras não seria possível na língua portuguesa.

Nathansohn e Brunet (2009) mostram claramente esse fato da negligência da história para com as mulheres ao citarem a história de Augusta Ada Byron, condessa de Lovelace, ou simplesmente Ada Lovelace. Filha da matemática Ann Isabella Milbanke e do poeta Lord Byron, é considerada a primeira programadora da história. Contudo, esse título surgiu graças a esforços feministas, pois ela sempre foi mais conhecida como por ter “ajudado” o inventor das primeiras calculadoras, Charles Babbage; “A informática tem um pai, mas sua mãe é ocultada” (ibid.; p.175).

Tais movimentos existem em prol de uma luta pela igualdade e para combater uma masculinização que ocorre no ciberespaço (URSUA, 2008). Isso pois, em pleno século XXI, as mulheres seguem sendo excluídas do acesso a esse ambiente (NATHANSOHN e BRUNET, 2009), e o que se vê é um reflexo do que ocorre no mundo real/físico, dominando pelo patriarcado e pelo machismo. Em seu artigo, Medina (2013) trabalha expressões sexistas na língua portuguesa utilizadas de forma machistas e preconceituosas pelo Blog Testosterona⁷, ele é um claro exemplo de como a internet engloba comportamentos arraigados na cultura patriarcal. “Percebe-se que esse pensamento está fortemente ligado a uma sociedade patriarcal, que oprime tudo aquilo que não se assemelha à classe dominante, ou seja, homens brancos e heterossexuais” (ibid., p. 58).

Nathansohn e Brunet (2009) levantam algumas das possíveis causas dessa desigualdade: letramento⁸, pois para acessar a rede é necessário saber ler e escrever; capacitação em informática básica e domínio de inglês; recursos para pagamento dos acessos; conteúdos úteis a mulheres; inserção de mulheres no desenvolvimento da ciência e da tecnologia.

Ainda assim, nos últimos doze anos é possível perceber uma mudança nesses fatos e o ciberespaço possibilitando um maior engajamento de mulheres interessadas nas pautas feministas, conforme Martinez (2019; p. 34-35):

Como a maioria dos movimentos sociais, o feminismo tem se valido de uma dinâmica reticular formada pelas plataformas de comunicação digital – blogs, redes sociais, vídeos, para difundir e promover suas pautas e reivindicações. Dentro dessa dinâmica, observa-se que nos últimos doze anos houve uma transformação em relação aos sujeitos desse debate, indicando a adesão cada vez mais precoce de jovens interessadas nas pautas feministas, o que está ligado diretamente ao fácil acesso e à familiarização desse

⁷ <https://www.testosterona.blog.br/>

⁸ Pesquisas recentes do IBGE mostram que, apesar de pouco percentual, o maior índice de analfabetismo no Brasil ainda incide sobre as mulheres, com 19,1% em detrimento de 18% de homens que não sabem ler. Disponível em: <<https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18317-educacao.html>>. Acesso em: 11 de nov. 2019.

público com as tecnologias. O crescente interesse pela temática feminista revela tanto o aumento dos espaços ocupados pelas feministas na internet quanto o extravasamento de suas preocupações para espaços da mídia internacional. Através da atualização contínua dos seus significados, essa nova “consciência feminista” tem sugerido uma nova política teleológica que congrega novas estratégias e formas de (re)produção do conhecimento.

Medina (2013) levanta uma questão presente no dia a dia que é prejudicial a essa dinâmica. Em seu trabalho, ela cita que mulheres e movimentos feministas são inimigos dessa masculinidade e tudo o mais que de alguma forma conteste sua dominância. Entretanto, ao elevar esses atores ao patamar de adversário, os possuidores dessa masculinidade esquecem que o responsável por criar o estereótipo de “macho dominante” não são esses movimentos, mas sim a sociedade com sua cultura, que dita desde o modo de vestir até o de se comportar.

Na seção seguinte será aprofundada a questão da existência ou não de um sexismo na língua e a influência do meio como determinação desse fenômeno.

2.1 Sexismo na língua

A análise de determinadas estruturas textuais e de discurso - estas que muitas vezes estão enraizadas na sociedade de forma preconceituosa, reafirmando preconceitos e modelos rígidos de identidade (MEDINA, 2013) - é essencial para discussão desse fenômeno do sexismo, que muito se ouve falar, estando presente até mesmo nas línguas. E isso vem se tornando alvo de discussões “no âmbito da linguística e da literatura pela Análise do Discurso Crítica e pela Crítica Cultural, que veem nas construções discursivas perigosas armadilhas para reafirmação de relações de poder assimétricas” (FOUCAULT, 1987 *apud*. MEDINA, 2013 p. 59).

Apesar de ainda se encontrar em alta a questão do preconceito e do sexismo, as mulheres já na década de 1940 questionavam o papel a elas relegado na sociedade de sempre serem femininas e atrativas aos homens (BEAUVOIR, 1970). Além disso, também estavam a conquistar independência, no sentido de ocuparem funções relevantes, o direito ao voto e, no pós Segunda Guerra Mundial, o de trabalhar fora. Ainda assim é ressaltado pela a autora (*ibid.*) que, tais avanços não eram vistos com bons olhos, pois a sociedade carecia de mudanças profundas, principalmente com relação à mentalidade.

Apesar desses avanços na década de 1940, os estudos sobre o sexismo na língua já levantavam interesse desde o século XVI. Entretanto, só com o avanço dos estudos linguísticos naquele mesmo século XX, é que eles ganhariam relevância. Esse impulso foi dado pelos movimentos feministas que começaram a divisar o machismo arraigado na língua, o que viria a

suscitar, de maneira até imperiosa, modificações ou mesmo uma reformulação linguística (BRITO, SOUSA e SILVA, 2015).

O século XX, por volta dos anos de 1960, trouxe consigo a segunda onda do movimento feminista (BRITO, SOUSA e SILVA, 2015). Impregnado no movimento, veio também o avanço de uma disciplina responsável pelo papel da significação do sexo e o papel sexual da linguagem (FOREL, 1983 *apud* BRITO, SOUSA e SILVA, 2015), enumerando três linhas de estudos embaladas pelo movimento feminista:

1. O estudo da utilização da linguagem, atualmente parte da sociolinguística;
2. A estrutura da língua, tida pelo movimento uma como figura machista;
3. “O estudo da história da língua e das possibilidades de planificá-la sistematicamente, ou seja, constatado o desnível entre a representação dos sexos na língua, esta deveria ser planificada para pôr termo à primazia do sexo masculino” (BRITO, SOUSA e SILVA, 2015, p. 1917).

2.1.1 Sexo X Gênero

Câmara Jr. (1999, p.88) fez a definição de gênero na língua como “uma distribuição. Em classes mórficas, para os nomes, da mesma sorte que o são as conjugações para os verbos”. O autor também declarou haver uma certa incoerência e confusão que as gramáticas tradicionais do português fazem a respeito da flexão de gênero.

A flexão de gênero é exposta de uma maneira incoerente e confusa nas gramáticas tradicionais do português. Em primeiro lugar, em virtude de uma incompreensão semântica da sua natureza. Costuma ser associada intimamente ao sexo dos seres. Ora, contra essa interpretação falam duas considerações fundamentais. Uma é que o gênero abrange todos os nomes substantivos portugueses, quer se refiram a seres animais, providos de sexo, quer designem apenas ‘coisas’, como casa, ponte, andaiá, femininos, ou palácio, pente, sofá, masculinos. [...]. Depois, mesmo em substantivos referentes a animais ou pessoas há discrepância entre gênero e sexo, não poucas vezes. Assim, testemunha é sempre feminino, quer se trate de homem ou mulher, e cônjuge, sempre masculino, aplica-se ao esposo e à esposa. Para os animais, temos os chamados substantivos epicenos, como cobra, sempre feminino, e tigre, sempre masculino. (CAMARA Jr.,1999, p. 88)

Sobre a explicação provida pela linguística da diferenciação de Sexo e Gênero, Brito, Sousa e Silva (2015) explicam que o sexo está pautado na biologia, ao passo que gênero é parte do campo linguístico. Forel (1983) vai além na explanação, declarando que ambos os termos não se misturam, ou seja, sexo e gênero não estariam ligados no plano da língua, uma vez que há muitas ocorrências onde ambos não coincidem e até se contradizem.

Brito, Sousa e Silva (2015) então ressaltam que o problema entre esses dois está nos casos em que a variável sexual recai sobre a língua, forçando uma flexão de gênero nas palavras. Os autores explanam os casos onde um elemento masculino, em meio a todo um universo feminino, basta para que todos aqueles citados sejam flexionados para o masculino. Exemplo deste parágrafo, onde uma das autoras é uma mulher, porém como os outros dois são homens, a flexão os transforma em “os autores”.

As possíveis relações entre sexo e gênero na língua são sintetizadas por Forel (1983, p. 26), por meio de um esquema de três tipos de séries, em que cada termo insere-se em uma série: a) sozinho, se não existir qualquer flexão possível, constituindo uma série de um termo (por exemplo, padre não conta com um substantivo correlato do gênero feminino); b) com mais de um termo, se for possível a flexão sem mudança semântica (o governante e a governanta, por exemplo, não integram a mesma série, pois não partilham a mesma essência semântica. Primo e prima, ao seu turno, integram a mesma série); e c) com mais dois termos, caso inexistente em português, em que há um termo masculino, outro feminino, e um que não marca qualquer gênero. (BRITO, SOUSA e SILVA, 2015, p. 1920).

A partir disso, os supracitados autores listam um esquema de séries analíticas para exemplificar a realidade da relação sexo e gênero na língua. O quadro a seguir, feito com exemplos dos mesmos autores (ibid., p. 1919-1920), exemplifica a proposta em relação aos termos.

Quadro 1 - Séries de exemplificação da relação sexo e gênero na língua

Séries	Exemplos
Séries de um termo	
Só masculino	Padre
Só feminino	Ama de leite
Indiferente	<i>They</i> (pronome da língua inglesa)
Séries de dois termos	
Só masculino	Primo
Só feminino	Prima
Só feminino	Francesas (termo marcado)
Só masculino ou masculino e feminino	Franceses (termo não marcado)
Só masculino	<i>Male Nurses</i> (termo marcado) (Substantivo da língua inglesa)
Só feminino ou feminino e masculino	<i>Nurses</i> (termo não marcado) (Substantivo da língua inglesa)
Série de três termos	
Só masculino	<i>vir</i> ⁹ (termo marcado) (substantivo da língua latina)
Só feminino	<i>mulier</i> (termo marcado) (substantivo da língua latina)
Indiferente	<i>homo</i> (termo não marcado) (substantivo da língua latina)

Fonte: próprio autor

⁹ *Vir*, é um substantivo latino do gênero masculino que significa “um”, em tradução literal. Podendo fazer alusão a “homem”, por exemplo. (Nota do autor)

Entendendo-se, segundo os autores, que “termos marcados” são aqueles exclusivos de um sexo, como *francesas*, que não existe hipótese de se referir a homens, seja apenas um ou mais. Já os “termos não marcados” são aqueles não possuem exclusividade de um sexo, franceses, pode tanto se referir a um grupo de homens nascidos na França, ou homens e mulheres do país.

A seção seguinte vai apresentar o Google Tradutor, bem como falar das dificuldades presentes para a tradução automática em fazer uma correspondência adequada entre o texto original e o resultado dessa TA. Além disso, levantar-se-á a questão do sexismo na língua também relacionada a tradução automática.

3. GOOGLE TRADUTOR

Esta seção, com explicações de o que é e como funciona o Google Tradutor, foi elaborada com base em um artigo do site informativo sobre tecnologias, Olhar Digital, escrito por Lucas Carvalho em 21 de dezembro de 2018.

De acordo com Carvalho (2018), o Google Tradutor é um aplicativo para *smartphones*, mas que também pode ser acessado por um navegador *web*¹⁰. Atualmente conta com tradução para mais de 103 idiomas distintos em seu sistema.

Seu funcionamento teve início ainda segundo o autor, em 26 de abril de 2006. Entretanto, nesse começo o aplicativo só contava com a tradução do idioma inglês para o árabe e vice-versa. Em detrimento do que ocorre atualmente com a expansão dessas mais de cem línguas, um montante superior a trinta milhões de sentenças por ano é traduzido no Google Tradutor.

Inicialmente, para elaborar as TAs, usava-se como base um sistema estatístico conhecido como SMT (*Statistical Machine Translation*¹¹). Isto é, toda tradução era feita primeiro para o inglês e só depois disso passava-se para a língua alvo, com base em documentos e livros presentes no acervo do Google. Por conseguinte, as traduções eram consideradas malfeitas, pois o sistema fazia um cruzamento de referências armazenadas em seu banco de dados, traduzindo palavras e expressões de maneira isolada e literal (CARVALHO, 2018).

Em 2016, a TA do Google passou a utilizar um sistema neural. Isso porque, mesmo com o uso contínuo do sistema estatístico, por mais que ele aprendesse, as TAs ainda permaneciam

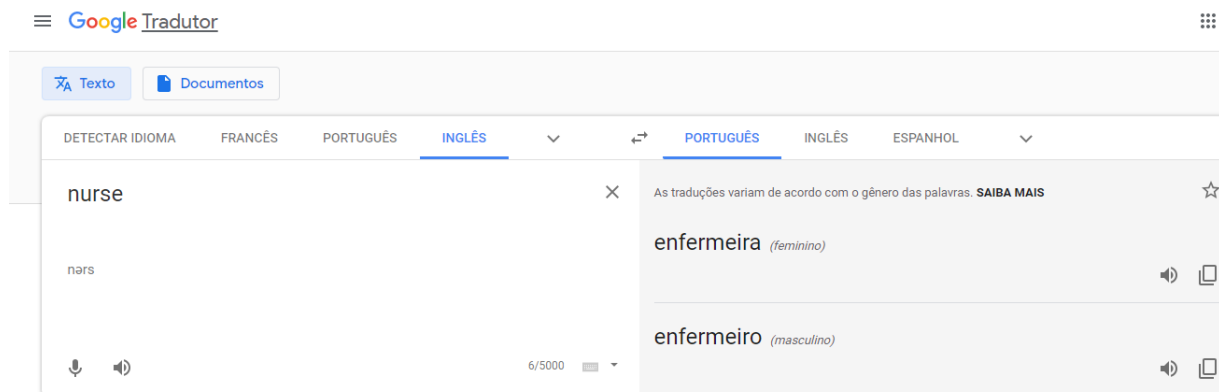
¹⁰ No endereço <<https://translate.google.com.br/>>

¹¹ Tradução Automática Estatística (nota do autor).

muito aquém do esperado. A partir de então, as sentenças passaram a ser processadas por inteligência artificial (IA), transpondo blocos inteiros de frases de uma vez só, não mais palavra por palavra e, com a ajuda da IA, o sistema foi aprendendo sozinho a realizar TAs mais precisas, por conseguir levar o contexto em consideração (CARVALHO, 2018).

No ano de 2018 (ibid.) foram adotados parâmetros para reduzir o viés de gênero. Expressões ou palavras que podem ter iguais significados, no masculino e no feminino, passaram a ter duas traduções, para que o uso de apenas uma não fosse imposto. Como é possível notar na figura abaixo:

Figura 7 - Tradução da palavra *Nurse* para o português



Fonte: captura de tela do Google Tradutor da tradução da palavra *nurse*, do inglês para o português

A captura de tela acima da utilização do Google Tradutor exemplifica o que é dito pelo autor, ao se traduzir simplesmente “nurse”, que no português pode ser tanto “enfermeiro”, como “enfermeira”, ele sugere as duas opções. E o próprio sistema explica sobre isso. Ao se clicar na opção “SAIBA MAIS”¹², ao lado da frase “As traduções variam de acordo com o gênero”, o Google diz, além da explicação do artigo, que a ordem em que as traduções surgem é baseada no sistema de ordem alfabética, assim tem-se a explicação do porquê a palavra no feminino vem primeiro, pois no português, por exemplo, “feminino” vem antes de “masculino” por ordem alfabética.

A próxima seção discutirá sobre o viés de gênero na tradução automática e como se espera que aconteça o alinhamento correto das sentenças de origem e de saída em uma TA.

¹² Disponível em <https://support.google.com/translate/answer/9179237?p=gendered_translations&hl=pt-BR&visit_id=637096059350205849-2043882184&rd=1>. Acesso em 17 de novembro de 2018.

3.1 Tradução automática e viés de gênero

Apesar de as primeiras ideias sobre tradução automática terem surgido no século XVII, com René Descartes propondo uma “língua universal”, foi só a partir de 1950 que elas começaram a se concretizar “com um memorando pioneiro de Warren Weaver [...] discutindo a possibilidade de empregar computadores para realizarem tradução automática” (PRATES, AVELAR e LAMB, 2018; p. 1, tradução do autor). Destarte, o famoso experimento Georgetown-IBM foi uma demonstração influente da tradução automática, realizada em 7 de janeiro de 1954. Desenvolvido em conjunto pela Universidade de Georgetown e pela IBM, o experimento envolveu a tradução totalmente automática de mais de sessenta frases do russo para o inglês (GORDIN, 2015).

Apesar de hoje em dia apresentarem um sucesso comercial enorme, as ferramentas de tradução automática possuem várias críticas. Noam Chomsky, fundador da linguística gerativista, cita que as realizações da tradução automática, embora sejam bem-sucedidas em um sentido particular, não são exitosas no sentido em que a ciência já se interessou: elas meramente fornecem maneiras eficazes, segundo Chomsky, de aproximar dados não analisados (Chomsky, 2011; Norvig, 2017 *apud*. Prates, Avelar e Lamb, 2018).

Várias vezes, o que se observa é uma dificuldade, após a realização de uma tradução automática envolvendo documentos intersemióticos de o texto-imagem, por exemplo, ficaram alinhados, entendendo-se por alinhamento o que definem Veltroni e Caseli (2018): uma tarefa de encontrar correspondentes específicos entre elementos textuais (palavras, expressões, etc.) e elementos visuais que os acompanham. Tradicionalmente, o alinhamento vêm utilizando textos paralelos para achar correspondentes entre textos fonte e alvo, sendo entendidos como textos paralelos aqueles escritos em uma língua acompanhados de suas traduções para outras línguas.

Como já supracitado, há um movimento recente, tanto do Google como de outros estudos para que esse alinhamento, ainda que primeiramente apenas sobre as TAs texto-texto, seja cada vez mais presente.

Recentemente, tem havido uma preocupação crescente na academia, nos laboratórios de pesquisa industrial e na mídia comercial convencional sobre o fenômeno chamado de viés de máquina, onde modelos estatísticos treinados (sem o conhecimento de seus criadores) crescem para refletir polêmicas assimetrias sociais, como gênero ou preconceito racial. Recentemente, foi sugerido que um número significativo de ferramentas de Inteligência Artificial tem uma influência prejudicial em relação a alguma minoria, com relatos de preditores racistas de comportamento criminoso, o iPhone X da Apple não consegue diferenciar entre dois povos asiáticos distintos e o agora infame caso de fotos

do Google classificarem erroneamente os negros como gorilas¹³ (PRATES, AVELAR E LAMB, 2018; p. 1).

O Google Tradutor, de fato mostra uma forte tendência à tradução errônea relacionada a profissões tipicamente desproporcionais na quantidade de homens e mulheres (como as da área da ciência, tecnologia, engenharia e matemática) e, mesmo que não se espere uma proporção de 50:50 na distribuição de gênero na tradução, o que o ele faz é traduzir para o masculino muito mais do que se espera, mesmo levando em conta dados demográficos (PRATES, AVELAR E LAMB, 2018).

Em um artigo de 2014, Londa Schiebinger sugeriu que a pesquisa científica falhou ao não levar questões de gênero em conta, argumentando que o fenômeno da padronização masculina nas novas tecnologias como o Google Tradutor, fornece uma janela para essa assimetria (SCHIEBINGER, 2014 *apud*. PRATES, AVELAR E LAMB, 2018).

A seguir será tratado de como ocorreu o levantamento de dados, a coleta e tratamento deles, bem como o objeto deste estudo.

4. METODOLOGIA

Para o desenvolvimento deste estudo foram coletados cinco pares de documentos multissemióticos. O banco de dados possui uma pequena extensão pelo fato de se prezar pelo caráter exploratório e qualitativo na pesquisa, uma vez que ela é uma continuação do banco de dados iniciado na pesquisa de Rodrigues e Pires (2019).

4.1 As Fontes de Dados

Conforme anteriormente mencionado, a pesquisa possui um caráter exploratório e qualitativo, uma vez que seu principal objetivo é explorar o achado da pesquisa feita anteriormente por Rodrigues e Pires (2019) para uma investigação mais precisa do fenômeno observado nos erros de TA em documentos multissemióticos com mulheres como figuras de destaque.

Todos os pares de imagens foram retirados do site de notícias *BBC News Online*¹⁴, de propriedade da *British Broadcasting Corporation*. É uma página web vinculada ao canal de

¹³ Reportagem disponível em: <<https://exame.abril.com.br/tecnologia/google-marca-fotos-de-casal-de-negros-como-gorilas/>>.

¹⁴ No endereço: < <https://www.bbc.com/>>

notícias *BBC Chanel*, canal da TV aberta britânica, que não só divulga notícias do Reino Unido, mas do mundo todo. Apesar de a página existir em várias línguas pelo mundo, este estudo se concentrou na página em inglês e em sua tradução para o português pelo tradutor automático, que são os objetos deste estudo.

A variedade de notícias da *BBC News Online* foi um dos principais fatores na escolha, por ela ser a base da fonte de dados desse estudo, uma vez que possibilitou uma exploração de diferentes tipos de notícias, bem como de reportagens com base em fatos advindos de todo o mundo. Finalmente, foi onde também os estudos base para este documento tiveram seus achados mais relevantes (vide Figuras 1 e 6).

4.2 Métodos e Ferramentas

Foram selecionados cinco pares de imagens da BBC para a realização deste estudo. A seleção se deu, primeiramente, buscando o endereço do site pelo navegador *web* Google Chrome, uma vez que ele já possui embutida a ferramenta principal de tradução automática aqui utilizada, o Google Tradutor, ferramenta de TA que, recentemente, incorporou tecnologia de tradução neural e tem um montante de mais de 200 milhões de usuários por dia (JOHNSON ET AL., 2017 apud PRATES, AVELAR E LAMB, 2018).

Os documentos selecionados necessitavam ser formados por múltiplos modos, além de as imagens apresentarem mulheres como figuras de destaque com um texto que, ao ser traduzido para o português gerasse uma ambiguidade e, conseqüentemente, a possibilidade de trabalhar e tratar da questão dos gêneros masculino e feminino, objetivo deste estudo. Desta feita, os modos utilizados, legenda *versus* imagem, apresentaram uma maior praticidade no descobrimento das incompatibilidades intersemióticas relacionadas ao gênero.

Apesar de não se tratar de um corpus, mas de um banco de dados, procedimentos de tratamento de corpora foram adotados para a etiquetagem, definida por Berber Sardinha (2004, p. 150) como “a inserção de informações referentes a cada unidade de texto (morfológica, sintática, semântica, discursiva). Por ser executada por computador, a etiquetagem automática permite o tratamento de grandes quantidades de texto rapidamente”. No entanto, essa definição de Berber Sardinha se aplica apenas a corpora monomodais, ou seja, aqueles que possuem um modo em sua formação. Uma vez que com corpora multimodais não existe uma padronização a ser adotada na etiquetagem, sendo este procedimento realizado de forma semi-

automática ou manual, pois além da falta de padrão, cada corpus multimodal possui suas peculiaridades e particularidades, ficando a cargo do pesquisador a definição do melhor esquema que se adapte para etiquetar (ABUCZKI; GHAZALEH, 2013).

Portando, para este estudo, adotou-se a ferramenta *Nimbus Capture*, uma ferramenta não especializada – que não foi desenvolvida para auxiliar o estudo de corpora, mas “que permite diferentes formas de etiquetagem” (ESPINDOLA, 2019; p. 20) - que é uma extensão² do navegador com a função principal de fazer capturas de tela, destacando-se ao oferecer esquemas de anotação e etiquetagem em documentos multimodais (ibid.).

As imagens coletadas e etiquetadas serão analisadas com base nos seguintes critérios na próxima seção: Qual incompatibilidade intersemiótica apresentam; qual desvio léxico-semântico; em qual série se encaixam e, por fim, onde está presente o sexismo no termo.

Na seção seguinte será feita a análise desses dados coletados, introduzindo inicialmente um panorama dos resultados, para depois esmiuçar cada um dos pares responsáveis pelos achados gerais.

5. ANÁLISE

Como enunciam Brito, Sousa e Silva (2015), uma das necessidades na emergência da luta por igualdade entre homens e mulheres é identificar onde está o sexismo e de que forma ele opera na sociedade. Um desses lugares identificado, com a segunda onda do feminismo, é a língua.

A língua por si, como observado nas seções anteriores, absorve muito dos comportamentos sexistas e vieses preconceituosos da cultura. Assim, isso será refletido em sistemas a ela inerentes, como os tradutores automáticos, que vão seguir essa tendência ainda que tenham uma inteligência artificial para evitar isso. Então, se em uma tradução de um documento composto apenas por texto apenas já existe um rompimento de relação do que se pretendia dizer no texto original com o texto de partida, no texto multissemiótico, a relação texto-imagem é mais evidentemente atacada com esse sexismo inerente.

Nos achados a serem discutidos detalhadamente a seguir, seguindo a tecnicidade da análise, em relação a textura intersemiótica de Liu & O’Halloran (2009), todos se encaixaram na maior das categorias - Relação Lógica ente Língua e Imagem, com a abordagem baseada em discurso, além de todas se encaixarem na categoria de relação intersemiótica de adição, uma

vez que nenhuma das imagens diz nada por si só, relacionadas ao contexto da reportagem. Por outro lado, diferente das manchetes que realmente trazem a notícia resumida em algumas linhas, porém quando elas se relacionam com a imagem, é que ocorre a quebra da textura, assim como em alguns casos de legendas das mesmas.

Em relação ao desvio léxico-semântico proposto por Vilar et al. (2006), todo o achado se concentrou em palavras incorretas, sentido e falha de desambiguação. Uma das possíveis razões disso é a neutralidade no inglês ser muito maior que a do português, geralmente se o pronome feminino não for bem marcado na sentença original de partida, a chegada vai quase sempre optar por uma tradução da frase no masculino.

No tocante às séries elencadas por Brito, Sousa e Silva (2015), as palavras traduzidas de forma errônea pelo Google Tradutor encaixam-se nas séries de dois termos, onde existe um termo para o masculino e outro para o feminino. Em relação aos trechos do original em inglês, na série de um termo, indiferente, ou seja, servindo tanto a um gênero quanto a outro, o que ratifica a dificuldade da TA na hora da desambiguação da tradução.

A seguir será feita a análise do banco de dados erigido na pesquisa com o objetivo de investigar nessas incompatibilidades intersemióticas geradas por erros de tradução automática a questão do gênero masculino e feminino e os impactos que tais erros afetam na questão do sexismo linguístico.

5.1 Qandeel Baloch

A notícia¹⁵ de onde foi extraída a imagem é sobre o assassinato de Qandeel Baloch, 26, em 2016. Estrela das mídias sociais, ela foi estrangulada pelos próprios irmãos, que alegaram a terem matado por ela ter trazido “desonra” à família. A manchete se refere ao fato de os pais deles, também dela, tentarem a todo custo liberarem os filhos, sem sucesso.

¹⁵ Disponível em: <<https://www.bbc.com/news/world-asia-49433097>>. Acesso em: 14 de setembro de 2019.

Figura 8 - Qandeel Baloch was killed by her brothers

Qandeel Baloch: Parents fail to free brothers accused of killing sister

1 hour ago

f [social media icons] Share



Figura 9 - Qandeel Baloch assassinada pelos irmãos

Qandeel Baloch: Pais não libertam irmãos acusados de matar irmã

1 hora atrás

f [social media icons] Compartilhar



- **Textura Intersemiótica quebrada:** relações lógicas entre língua e imagem – abordagem baseada em discurso - relação intersemiótica de adição.
- **Desvio Léxico-semântico:** palavra incorreta – sentido – falha de desambiguação.
- **Série de termos da palavra incorreta:** Série de dois termos (morto/morta).
- **De que forma é sexista?**

Killed, no documento original, poderia ser classificada como Série de um termo, indiferente, de acordo com a classificação de Brito, Sousa e Silva (2015). Ou seja, na língua inglesa, ele pode fazer referências tanto a uma mulher como um homem que vai funcionar bem. Porém, o sistema da TA optou pela forma masculina em primeiro lugar, mesmo a etiquetagem mostrando que era uma mulher, uma “irmã”, mas a legenda optou por dizer que foi “morto”. Como já supracitado, o sistema pode reproduzir comportamentos e padrões do mundo real, um deles é eleger o masculino em primeiro lugar, como aconteceu com esse termo.

5.2 Lyubov Sobol

Essa reportagem¹⁶ possui um vídeo e um texto introdutório na lateral. É sobre Lyubov Sobol, uma grande influenciadora digital da Rússia, ativista contra a corrupção, que movimentou imensos protestos contra o governo do presidente Vladimir Putin ao longo de uma década. Ela é conhecida como “invisível”, pois é proibida de aparecer em mídias tradicionais, assim como se candidatar a cargos em eleições locais. Mesmo assim, apresenta essa força motriz de arrastar multidões a protestos, palestras, tendo bastante destaque nos meios digitais de comunicação.

¹⁶ Disponível em <<https://www.bbc.com/news/av/entertainment-arts-50188580/lyubov-sobol-meet-russia-s-invisible-influencer>>. Acesso em 26 de out. de 2019.

Figura 10 - Lyubov Sobol, *the invisible influencer*

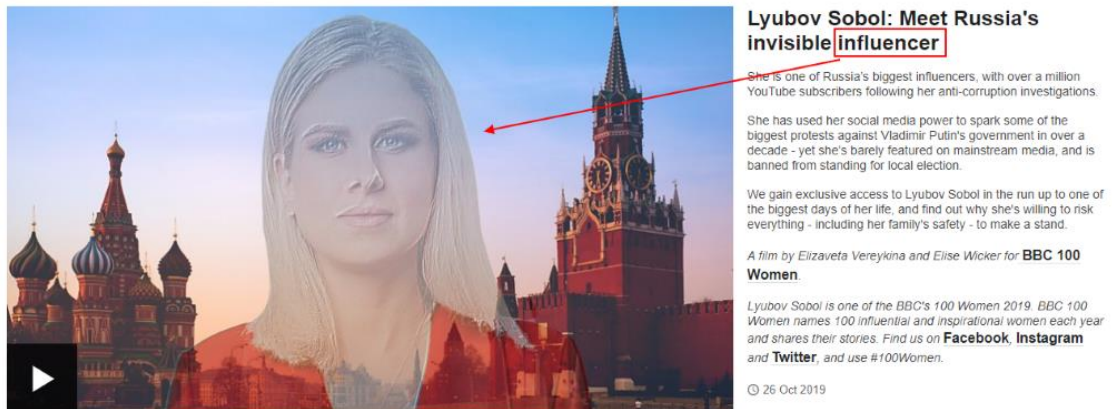
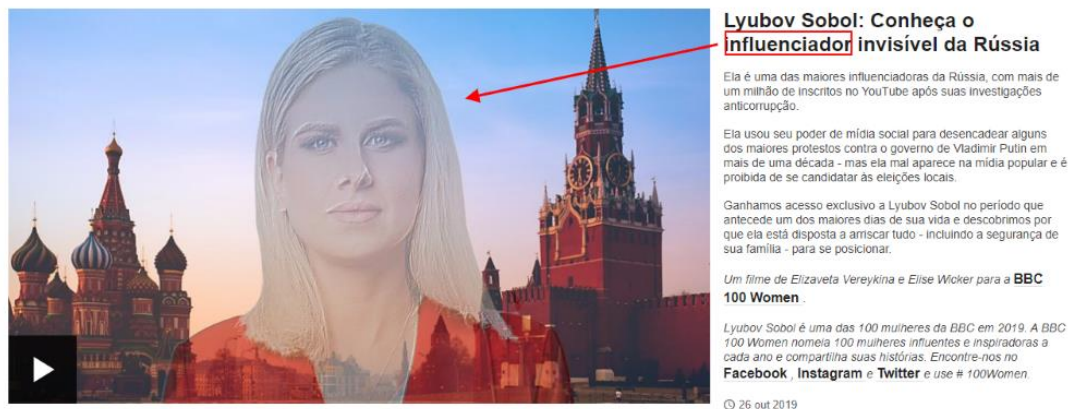


Figura 11 - Lyubov Sobol, a influenciadora invisível



- **Textura Intersemiótica quebrada:** relações lógicas entre língua e imagem – abordagem baseada em discurso - relação intersemiótica de adição.
- **Desvio Léxico-semântico:** palavra incorreta – sentido – falha de desambiguação.
- Série de termos da palavra incorreta:** Série de dois termos (influenciadora/influenciador).
- **De que forma é sexista?**

Influencer, assim como nas imagens analisadas anteriormente, poderia ser classificada como Série de um Termo, Indiferente (BRITO, SOUSA e SILVA, 2015). O termo na língua inglesa pode aludir tanto a uma mulher como a um homem. Contudo, a TA optou pela forma masculina, ainda que, como mostra a etiquetagem, a manchete da notícia faça referência a uma pessoa do sexo feminino, que se destaca mesmo na foto translúcida, para reforçar a textura intersemiótica que faz alusão ao título.

Segundo Milanezi e Darci (2019, p. 280), ferramentas criadas na era da internet como, por exemplo, “os blogs, o Twitter, o YouTube e o Instagram [...] colocaram um número incontável de produtores de conteúdo independentes de uma mídia hegemônica em um patamar de visibilidade como nunca antes”. Tal citação reforça o que foi levantado nos primeiros capítulos deste estudo, onde se falou sobre um ambiente que, mesmo sendo construído recentemente, como a internet, levanta a questão do preconceito e do viés de gênero. Uma mulher não pode ser uma influenciadora, só pode ser influenciada, como há muitos e muitos anos vêm ocorrendo.

Além disso, a expressão *digital influencer* é bastante recente, e muitas vezes a tendência no português brasileiro, principalmente aqueles falado pelos mais jovens, é absorver essas expressões em alta, em vez de aportuguesá-las ou achar um correspondente na língua de chegada. Mas de fato, com esse termo, um pouco dos dois ocorre. Muito se vê hoje em dia a utilização sem tradução, como as figuras abaixo, extraídas de manchetes de notícias, podem demonstrar:

Figura 12 - Diversidade cultural com digital influencer



Fonte: <<https://tvefamosos.uol.com.br/colunas/flavio-ricco/2019/11/18/rede-tv-estrelia-atracao-sobre-diversidade-cultural-com-digital-influencer.htm>>. Acesso em 18 de nov. 2019.

Figura 13 - Instagram esconde *likes* dos *digital influencers*



Fonte: <https://gazetaweb.globo.com/porta1/noticia/2019/11/esconder-likes-instagram-pondera-alternativa-apos-polemicas_90715.php>. Acesso em: 18 de nov. 2019.

Figura 14 - Ascensão da Digital Influencer Júlia Lawrence

PUBLICIDADE CORPORATIVA

A ascensão do digital influencer, Julia Lawrence é um símbolo da nova forma de se fazer publicidade

A modelo e digital influencer Julia Lawrence representa uma tendência crescente da segunda década do século XXI, mostrando a publicidade digital como solução para uma abordagem mais próxima do cliente

Por **JayPRO**
 © 18 out 2019, 18h25- Atualizado em 1 nov 2019, 15h12



Fonte: <<https://veja.abril.com.br/economia/jp/a-ascensao-do-digital-influencer-julia-lawrence-e-um-simbolo-da-nova-forma-de-se-fazer-publicidade/>>. Acesso em: 18 de nov. 2019.

Todas as capturas de tela acima foram retiradas de notícias sobre influenciadores digitais, em todas o que se percebe é uma agregação do termo em inglês no português, sem nenhum destaque a essa expressão estrangeira, ou seja, a tratando já como parte do vocabulário.

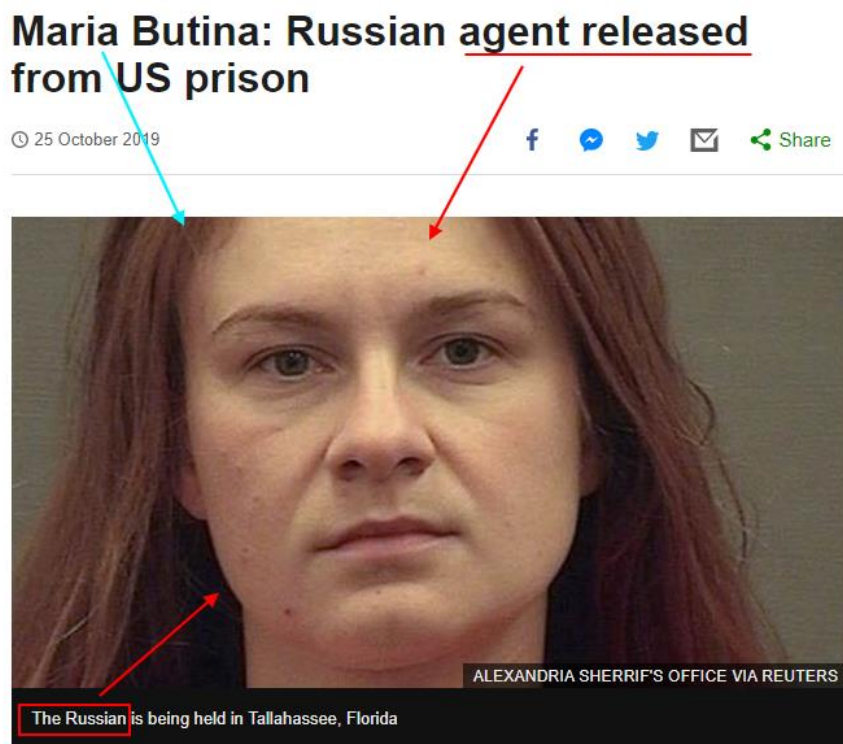
É na figura 14, ainda se pode notar uma incompatibilidade intersemiótica, etiquetada da mesma maneira que as outras (vide seção “metodologia”), onde por um erro textual, o viés de gênero é reforçado, uma vez que o termo absorvido pela língua portuguesa cairia na Série de um Termo, Indiferente (BRITO, SOUSA e SILVA, 2015), porém a preposição “do” (de + o),

optou pelo artigo masculino “o”, fundido a preposição “de”, gerando assim uma incompatibilidade intersemiótica com viés de gênero, mesmo a matéria tendo sido produzida em língua portuguesa e não traduzida.

5.3 Maria Butina

A seguinte reportagem¹⁷ é a respeito de Maria Butina, uma espiã russa presa nos Estados Unidos e sentenciada a dezoito meses de prisão após ser pega.

Figura 15 - Russian *agent*, Maria Butina



¹⁷ Disponível em < <https://www.behindthename.com/name/mari> >. Acesso em 26 de out. 2019.

Figura 16 - Agente Russa, Maria Butina



– **Textura Intersemiótica quebrada:** relações lógicas entre língua e imagem – abordagem baseada em discurso - relação intersemiótica de adição.

– **Desvio Léxico-semântico:** palavra incorreta – sentido – falha de desambiguação.

– **Série de termos das palavras incorretas:** Série de dois termos (russa/russo e libertada/libertado).

– **De que forma é sexista?**

O problema aqui vem de vários ângulos. A relação intersemiótica é quebrada tanto na legenda como na manchete da imagem. *Russian* e *released*, ambas palavras que também figuram como neutras no inglês, servindo ao feminino e ao masculino. Porém, na imagem que informa a situação da agente russa Maria Butina, elas se tornaram, respectivamente, “russo” (por duas vezes) e “libertado”.

A etiqueta formada pela seta azul representa a marcação de Maria e a figura na imagem. Maria, por senso comum, é um nome do gênero feminino, adotado não só no português – como é possível de se ver na imagem, também no russo – e em vários outros idiomas, como no próprio inglês. Apesar disso, ele também é utilizado como nome masculino, mas na maioria das vezes

como o segundo nome de um prenome composto, como José Maria, ou Carlos Maria. O site da *web Behind the name*¹⁸, responsável por traçar origens de nomes e explicar seus significados, também confirma tal informação a respeito desse nome nos idiomas polonês e alemão. Além disso, ele deriva do hebreu (מָרְיָם) sendo utilizado muito antes da famosa mãe de Jesus Cristo.

Porém, no caso das figuras em análise, não resta dúvidas que Maria é a mulher em destaque, a agente russa que foi presa. “Agente”, é uma palavra que também se destaca, pois se encaixaria na série de um termo e indiferente. Caso não esteja marcada, o seu correlato em inglês, *agent*, o mesmo. Entretanto, é mais uma profissão que, apesar de historicamente sempre ter tido mulheres atuando, é associada a homens, e mesmo com o nome Maria, a TA optou por colocar o que podia no masculino.

5.4 Samantha Kureya “Gonyeti”

A seguinte notícia¹⁹ trata de Samantha Kureya, uma comediantes zimbabuense conhecida pelo nome artístico de Gonyeti, que foi sequestrada por homens mascarados da própria casa e brutalmente espancada e violentada. Há suspeitas de que o ocorrido foi por suas piadas críticas ao governo.

Figura 17 - Samantha Kureya, comedian from Zimbabwe

Zimbabwean comedian Gonyeti 'abducted and beaten' in Harare

© 22 August 2019

f WhatsApp Twitter Email Share



¹⁸ Disponível em: < <https://www.behindthename.com/name/maria> >. Acesso em: 24 nov. de 2019.

¹⁹ Disponível em: < <https://www.bbc.com/news/world-africa-49433387> >. Acesso em: 22 ago. 2019.

Figura 18 - Samantha Kureya, comediante do Zimbábue

Comediante zimbabuense Gonyeti 'seqüestrado e espancado' em Harare

3 horas atrás

f WhatsApp Twitter Email Compartilhar



- **Textura Intersemiótica quebrada:** relações lógicas entre língua e imagem – abordagem baseada em discurso - relação intersemiótica de adição.
- **Desvio Léxico-semântico:** palavra incorreta – sentido – falha de desambiguação.
- **Série de termos das palavras incorretas:** Série de dois termos (sequestrada/sequestrado e espancada/espancado).
- **De que forma é sexista?**

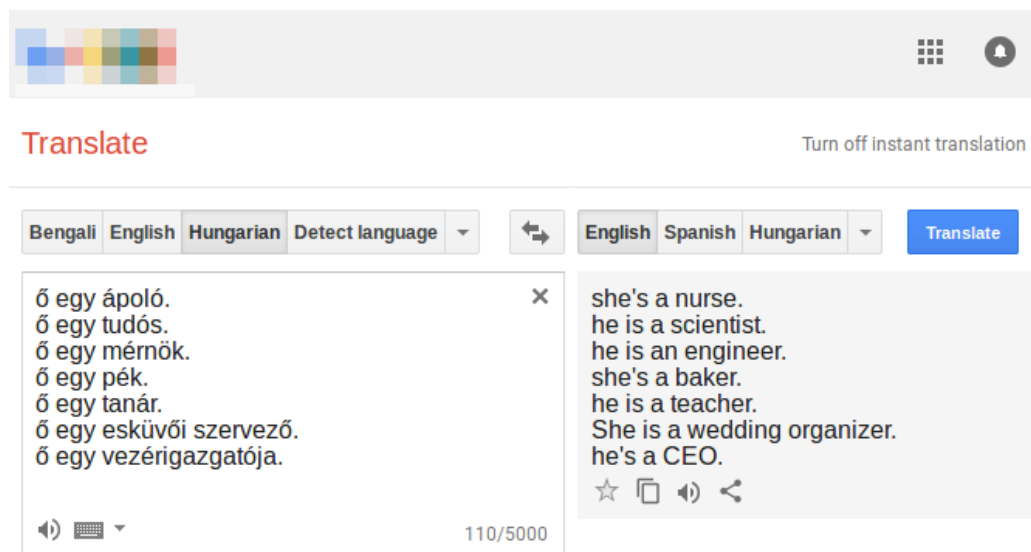
Mais uma vez, problemas de flexão dos adjetivos que, na língua inglesa, são neutros e na portuguesa, variam de acordo com o gênero. *Abducted* e *beaten*, em português, no contexto, o certo seria “sequestrada” e “espancada”, respectivamente, porém a TA optou pela forma masculina.

A etiquetagem feita em formato de caixas azuis na legenda da imagem destacam o pronome *she*, no inglês, “ela”, para exaltar que somente quando claramente está definido que é uma mulher a quem o texto se refere, a textura intersemiótica não é quebrada. E provavelmente a palavra *banned*, seria transformada em “proibido” ou “banido” em português.

Algo que poderia ser levantado aqui, novamente, seria a questão das profissões. As pessoas não estão acostumadas com uma mulher comediante. É mais um campo dominado pelos homens.

Para exemplificação desse fenômeno de viés de gênero proposto por Prates, Avelar e Lamb (2018), a figura abaixo, extraída do trabalho dos autores, mostra esse direcionamento da tradução da máquina, ainda que se parta de um idioma com maior neutralidade como o húngaro - uma vez que seu pronome pessoal da terceira pessoa, *Ő*, não distingue masculino e feminino - e a chegada seja outro idioma de certa neutralidade, como o inglês.

Figura 19 - Sentenças do Húngaro para o Inglês

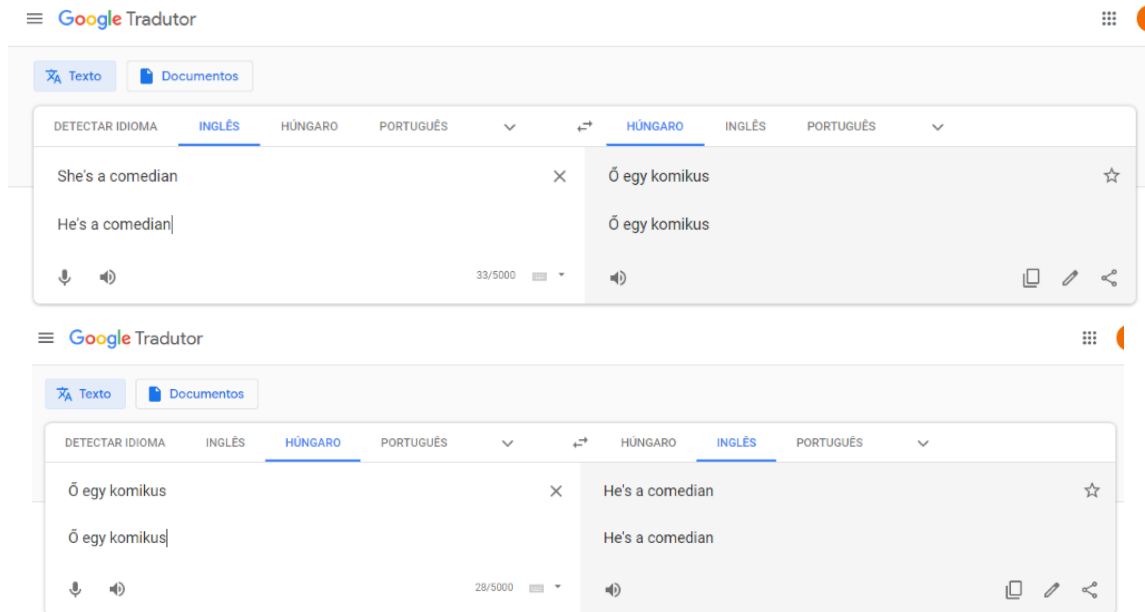


Fonte: Prates, Avelar e Lamb (2018, p. 4)

A captura de tela feita pelos autores, do Google Tradutor, mostra a questão de mesmo quando em um idioma neutro, profissões tradicionalmente associadas ao gênero masculino – como academia, engenharia e diretoria executiva – são interpretadas como sendo exercidas por homens, ao passo que a profissão *nurse* (enfermeira/o), mesmo sendo algo neutro em inglês, como já discutido por Brito, Sousa e Silva (2015) é interpretada como uma profissão feminina.

Seguindo a ideia dos autores, aplicadas a profissão de comediante, teríamos o seguinte:

Figura 20 - Capturas de tela do Google Tradutor



Fonte: capturas de tela do Google Tradutor.

As figuras acima são capturas de tela do Google Tradutor, utilizando novamente a tradução do inglês para o húngaro e vice-versa. Como é possível de se observar, *She's* ou *He's a comedian* (ela ou ele é um comediante), traduzidos para o outro idioma, se tornam neutros também resultando em *Ő egy komikus*. Porém, quando a tradução é invertida, mesmo sem a necessidade de digitar novamente os termos, pois clicando-se no botão das duas setas entre as línguas, os idiomas mudam de lugar como entrada e saída, assim como os que está previamente escrito nas caixas abaixo. E tendo *Ő egy komikus* como entrada, as duas traduções são para o masculino, *he's a comedian* (ele é um comediante).

5.5 Claus von Bülow

A seguinte reportagem⁶ é sobre a morte Claus von Bülow, um socialite que foi acusado do assassinato da esposa em 1982, e morreu aos 92 anos.

Figura 21 - Claus von Bülow, Socialite

Claus von Bülow: Socialite cleared of trying to murder his wife dies aged 92

© 31 May 2019

f     Share



Figura 22 - Claus von Bülow, Socialite

Claus von Bülow: Socialite liberada de tentar matar sua esposa morre aos 92 anos

© 31 maio 2019

f     Compartilhar



– **Textura Intersemiótica quebrada:** relações lógicas entre língua e imagem – abordagem baseada em discurso - relação intersemiótica de adição.

– **Desvio Léxico-semântico:** palavra incorreta – sentido – falha de desambiguação.

– **Série de termos da palavra incorreta:** Série de dois termos (liberado/liberada). Aqui cabe mencionar também socialite, uma vez que ela é a responsável pela falha na desambiguação. Sendo essa palavra parte da série de um termo, indiferente, no inglês e no português.

– **De que forma é sexista?**

Cleared, é um adjetivo, também neutro, que deveria ter sido traduzido por “liberado”, dado o contexto, pois o socialite em questão é Von Bülow, um homem. O sexismo, neste caso, pode ser observado pela tradução da palavra “socialite”.

A única tradução que causou incompatibilidade intersemiótica traduzindo para o feminino, quando deveria usar para o masculino, foi com a palavra em questão. Isso porque, como já dito, a TA costuma absorver comportamentos e vieses da sociedade, e “socialite” é geralmente associada a mulheres e, de certa forma, com um caráter pejorativo. Em particular, mulheres fúteis que nada mais fazem da vida além de serem “mulheres ricas”.

O dicionário da língua inglesa Macmillan tem a seguinte definição para socialite “alguém que vai a eventos sociais organizados por pessoas da alta sociedade”²⁰. A palavra “alguém” em destaque é proposital, justamente para exemplificar que tanto em inglês, como em português esse é um adjetivo de dois gêneros, ou na classificação de Brito, Sousa e Silva (2015), série de um termo e indiferente. Mesmo assim, nessa única palavra, a tradução foi feita de forma errônea. Reforçando esse estigma, as figuras em exemplo abaixo, onde a tradução está correta, a textura intersemiótica não foi quebrada, mas a pergunta que resta é, porque a única tradução correta foi com o emprego dessa palavra, por vezes estigmatizada e com viés.

²⁰ No original “someone who goes to social events organized by fashionable people”. Tradução do autor. Disponível em: < <https://www.macmillandictionary.com/dictionary/british/socialite>>. Acesso em: 12 nov. de 2019

Figura 23 - Ksenia Sobchak, Socialite

Russia socialite Ksenia Sobchak declares presidential bid

© 18 October 2017

f    Share

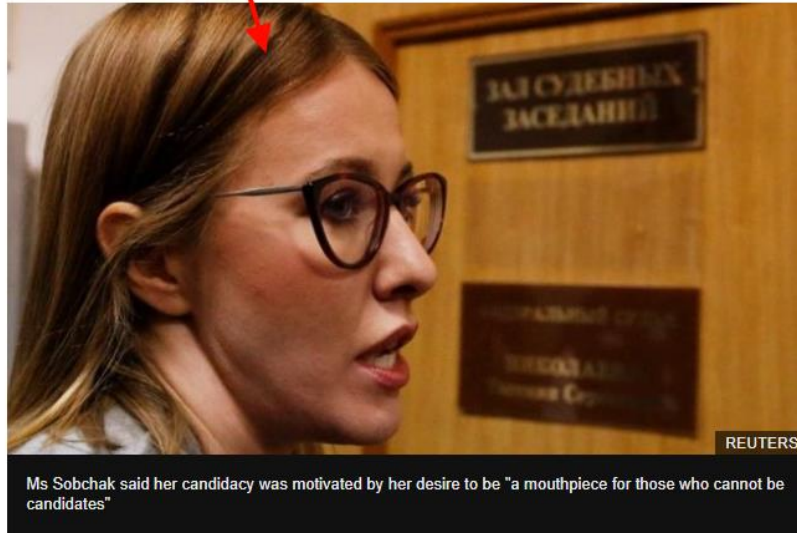


Figura 24 - Ksenia Sobchak, Socialite

Socialite russa Ksenia Sobchak declara candidatura presidencial

© 18 de outubro de 2017

f    Compartilhar



Fonte: < <https://www.bbc.com/news/world-europe-41669676> >. Acesso em: 19 nov. 2019.

Os dados aqui discutidos mostram claramente que padrões e vieses, além de preconceitos de gênero sofrido por mulheres na sociedade, são absorvidos por sistemas que foram criados para facilitar a vida de um usuário no dia a dia. O que levanta uma outra questão a ser discutida futuramente, o porquê disso? E o sexismo realmente estaria no sistema de TA, ou nas pessoas que o produziram?

Na próxima seção, apresentam-se as considerações finais, algumas dessas questões inicialmente propostas serão respondidas, além do levantamento de novas hipóteses para pesquisas futuras e a continuação do desenvolvimento do tema.

6. CONSIDERAÇÕES FINAIS

Neste estudo se mostrou um certo viés do tradutor automático, com base em problemas de traduções automáticas em documentos multimodais. Analisando um pequeno banco de dados de forma qualitativa e exploratória, foi possível perceber a discrepância em gênero masculino e feminino nos erros gerados por uma TA.

Esses erros se mostram relevantes, primeiramente por expor uma desigualdade presente na sociedade, que muitas vezes é vista como “pouca coisa” se comparada a “assuntos mais importantes”. Porém, a importância de se lutar por igualdade e direitos é constante, e é necessário de se levantar essas bandeiras quando são ignoradas. Além disso, um olhar que pode ser uma tentativa de desconstrução da masculinidade dominante na sociedade, que subjuga não só as mulheres, mas todos os que fogem do padrão de ser homem, hétero, branco e cisgênero.

Os resultados mostraram uma discrepância visível no comportamento do Google Tradutor ao fazer as traduções automáticas, apresentando erros ou não., quase sempre optando pelo gênero masculino em vez do feminino. E um único caso que deu preferência ao feminino, foi em uma palavra por vezes estereotipada e depreciativa como “socialite”. Essa suspeita se confirma como o fato de que, mesmo em línguas, a princípio, com uma maior neutralidade de gênero como o inglês, existe esse sexismo velado da máquina.

O objetivo principal deste estudo foi levantar e mostrar que esse problema existe e carece, e muito, de mais desenvolvimento, cumprido com êxito. Com relação às problemáticas levantadas, as mulheres antes não tinham voz, precisavam sempre aceitar ficarem relegadas a segundo plano como um objeto, mas isso mudou com as ondas do movimento feminista, que cresceu e hoje busca igualdade. Na língua, com a segunda onda do movimento feminista elas identificaram o problema e passaram a mostrá-lo e combater o sexismo que por vezes está ali presente de forma que geralmente não se percebe.

Em relação à cultura machista ser refletida no tradutor automático, é um fato. Pode-se, assim, dizer que sexista não é o tradutor, ou a língua, mas sim aqueles que desenvolvem esses softwares, que os criam e cuidam deles. Uma inteligência artificial com comportamentos machistas e racistas é só um espelho da sociedade onde está inserida, assim como a língua e suas regras ditadas por aqueles que se dizem dominantes e “machos”.

Uma reflexão que pode ser deixada para o futuro é: em quantas das línguas registradas no tradutor, mais de 130, esse fenômeno se repete? Alguns testes já feitos em línguas com maior diferenciação de gênero, similares ao português, como o espanhol e a francesa, mostram que tal fenômeno se repete. Com isso, conclui-se ser preciso aventar possibilidades de como seria

possível ensinar a máquina de maneira distinta, para que possa evitar esses comportamentos sociais que perpetuam a disparidade entre, não só, homens e mulheres, mas todos aqueles grupos historicamente sujeitos à preconceito e segregação perante a sociedade.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABUCZKI, A.; GHAZALEH, E.B. **An overview of multimodal corpora, annotation tool and schemes.** In: *Argumentum* 9, p. 86-98. Debreceni Egyetemi Kiadó, 2013.
- BATEMAN, J. A. **Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents.** New York: Palgrave MacMillan, 2008.
- BERBER SARDINHA, T. **Linguística de Corpus.** Editora Manole Ltda. São Paulo, 2004.
- BEAUVOIR, Simone. A mulher independente. In: **O segundo sexo II: a experiência vivida.** Tradução Sérgio Millet. Rio de Janeiro: Difusão Europeia do Livro, 1970.
- BRITO, D.T; SOUSA, V. V. e SILVA, J. A. A. da. **Sexismo na língua portuguesa.** *Revista Philologus*, Rio de Janeiro, n. 63, p. 1915-1928, setembro/dezembro, 2015. Disponível em: <<http://www.filologia.org.br/rph/63supl.html>>. Acesso em: 11 nov. 2019.
- CAMARA Jr., Joaquim Mattoso. **Estrutura da língua portuguesa.** 30. Ed. Petrópolis: Vozes, 1999.
- CARVALHO, Luís. **Google Tradutor: como funciona, dicas e tudo o que você precisa saber.** *Olhar Digital.* Disponível em: <<https://olhardigital.com.br/noticia/google-tradutor-como-funciona-dicas-e-tudo-o-que-voce-precisa-saber/80804>>. Acesso em: 19 nov. 2019.
- DARCIE, M.; MILANEZI, M. J. **“Blogueirinha rica não entende de Brasil”: cobrança pelo posicionamento político e manifestações de influenciadores na internet.** *Revista Extra-prensa*, v. 12, p. 274-289, 17 out. 2019.
- ESPINDOLA, Augusto Velloso. **Incompatibilidades intersemióticas geradas por tradução automática: um estudo exploratório sobre coleta, etiquetagem e anotação em corpus de documentos multimodais estáticos.** Monografia – Brasília: UnB, 2019. No Prelo.
- FOREL, Claire. Francesas, franceses... In: AEBISCHER, Vera; FOREL, Claire. (Orgs.). **Falas masculinas, falas femininas?** P. 23-34. São Paulo: Brasiliense, 1983.
- GORDIN, M.D. **Scientific Babel: how science was done before and after global English.** University of Chicago Press, Chicago, 2015.
- HUTCHINS, W. J. **Machine translation: A concise history.** *Journal of Translation Studies*, v.13,n.1-2, p. 29-70, 2010.
- KRESS, G.; VAN LEEUWEN, T. **Reading Images: The Grammar of Visual Design.** 2nd. ed. London: Routledge, 2006.
- LIU, Y.; O’HALLORAN, K. L. **Intersemiotic Texture: analyzing cohesive devices between language and images.** *Social Semiotics*, v. 19, n. 4, p. 367-388, 2009.
- MARTINEZ, Fabiana. **Feminismos em movimento no ciberespaço.** *Cad. Pagu*, Campinas , n. 56, e195612, 2019 . Disponível em <http://www.scielo.br/scielo.php?script=sci_art-text&pid=S0104-83332019000200502&lng=en&nrm=iso>. Acesso em 11 nov. 2019.

MEDINA, Marianna Michelle. **Configurações do discurso machista na internet: O blog Testosterona e os embates ideológicos na era digital**. Jangada: crítica | literatura | artes. Viçosa-MG, n.2, p. 57-74, julho/dezembro, 2013.

MILANEZI, M. J. F.; DARCIE, M. P. **“Blogueirinha rica não entende de Brasil”: cobrança pelo posicionamento político e manifestações de influenciadores na internet**. Extraprensa, São Paulo, v. 12, n. esp., p. 274 – 289, set. 2019.

NATHANSOHN, Leonor Graciela; BRUNET, Karla Schuch. **Ciberespacio y Mujeres, una tierra en transe**. Revista Latinoamericana de Ciencias de la comunicación, <<https://www.alaic.org/revista/index.php/alaic/index>>, n. 11, p. 170-181, julho/dezembro, 2009.

PIRES, T. B. **Ampliando olhares sobre a tradução automática online: um estudo exploratório de categorias de erros de máquina de tradução gerados em documentos multimodais**. Tese de doutorado—Brasília: UnB, 2017.

PRATES, M.; AVELAR, P. H. da C.; LAMB, L. C. **Assessing Gender Bias in Machine Translation: A Case Study with Google Translate**. Neural Comput & Applic. Rio Grande do sul: UFRS, setembro, 2018.

PLAZA, Júlio. **Tradução intersemiótica**. 1ª edição; 2ª reimpressão. São Paulo: Perspectiva, 2003.

RODRIGUES, Jeferson Viegas; PIRES, Thiago Blanch. **Catálogo de novas configurações intersemióticas geradas por resultados de tradução automática em documentos multimodais**. In: 25º Congresso de Iniciação Científica da Universidade de Brasília e 16º Congresso de Iniciação Científica do Distrito Federal. Seminário do Programa Institucional de Bolsas de Iniciação Científica do CNPq/ PIBIC, Brasília: Instituto de Letras, Departamento de Línguas Estrangeiras e Tradução da Universidade de Brasília - UnB, 2019.

URSUA, Nicanor. **La(s) identidad(es) en el ciberespacio. Una reflexión sobre la construcción de las identidades en la red (“online Identity”)**. Ontology Studies, País Basco n.8, p. 277-296, 2008.

VELTRONI, W. C.; CASELI, H. de M. **Text-Image Alignment in Portuguese News Using LinkPICS**. Computational Processing of the Portuguese Language: 13th International Conference, PROPOR 2018, Canela, Brasil, September 24–26, 2018. São Carlos: UFSCar P. 125-135, 2018.

VILAR, D. et al. **Error analysis of statistical machine translation output**. In: PROCEEDINGS OF LREC. Genoa: 2006.