

GUILHERME, Maurício Rubens de Carvalho; WALDOLATO, Rafaela Amaral. Um estudo da variação lexical dos itens "Índio" e "Indígena" em textos escritos de Língua Portuguesa. *Revista Intercâmbio*, v.LVII, 68908, 2025. São Paulo: LAEL/PUCSP. ISSN 2237-759X

<https://doi.org/10.23925/2237-759X2025V57e68908>

UM ESTUDO DA VARIAÇÃO LEXICAL DOS ITENS "ÍNDIO" E "INDÍGENA" EM TEXTOS ESCRITOS DE LÍNGUA PORTUGUESA

A STUDY OF THE LEXICAL VARIATION OF THE TERMS 'ÍNDIO' AND 'INDÍGENA' IN WRITTEN PORTUGUESE TEXTS

Maurício Rubens de Carvalho GUILHERME
(Universidade do Estado de Minas Gerais)
mauricio.guilherme@uemg.br¹

Rafaela Amaral WALDOLATO
(Universidade do Estado de Minas Gerais)
rafaelawaldolato@gmail.com²

RESUMO: O presente trabalho tem como objetivo descrever a variação lexical entre os itens "Índio" e "Indígena" em textos escritos de língua portuguesa a fim de saber como os falantes de língua portuguesa se referem aos povos originários e seus descendentes do Brasil. A pesquisa foi realizada por meio de dados presentes no site www.corpusdoportugues.org. Como resultado verificou-se que, embora alguns fatores favoreçam a ocorrência de "Indígena" como núcleo do sintagma nominal, o resultado não é significativo para indicar uma tendência de mudança em curso no período analisado.

PALAVRAS-CHAVE: Sociolinguística; variação lexical; índio; indígena.

ABSTRACT: *The present work aims to describe the lexical variation between the terms "Índio" and "Indígena" in written Portuguese texts in order to understand how Portuguese speakers refer to the indigenous peoples and their descendants in Brazil. The research was conducted using data available on the website www.corpusdoportugues.org. The results showed that, although some factors promote the occurrence of "Indígena" as the head of the noun phrase, the result is not significant enough to indicate a trend of change occurring during the analyzed period.*

KEYWORDS: *Sociolinguistics; lexical variation; indian; indigenous.*

¹ Doutor em Estudos Linguísticos pela Universidade Federal de Minas Gerais.

² Graduada em Letras pela Universidade do Estado de Minas Gerais.

1. Introdução

À luz da teoria Sociolinguística Variacionista, que segundo Cezário e Votre (2008) considera a língua como um fenômeno essencialmente social e evidencia os fatores diversidade e variabilidade como inerentes ao sistema linguístico, o presente trabalho tem como finalidade descrever a variação linguística entre os itens lexicais “Índio” e “Indígena” – utilizados para designar o mesmo referente – em situações reais de uso.

Ferreira e Lacerda (2021) destacam que o termo “Índio”, imposto aos povos originários desde a chegada de Cristóvão Colombo, o qual acreditava ter chegado às Índias, é considerado um termo pejorativo que “coloca a população não só como ‘diferente’ e alheia à sociedade contemporânea, como, também, a identifica como ‘atrasada’”. Ademais, Valquíria Kyalonã, do povo Xukuru da Serra de Ororubá em Pernambuco, enfatiza que esse item lexical desconsidera os povos que já habitavam o território – hoje chamado de Brasil – antes da colonização, reduzindo a pluralidade étnica e a diversidade cultural do país.

Em face das diferentes origens das palavras “Índio” e “Indígena” e do viés ideológico³, cultural e sócio-histórico presente nas escolhas lexicais desses itens, “a adoção do termo ‘indígena’ que significa ‘natural do lugar [sic] que se habita’ tem sido indicada como definição mais adequada para se referir aos povos originários” (FERREIRA; LACERDA 2021).

Dessa forma, analisar os grupos de fatores que propiciam a variação entre os itens lexicais “Índio” e “Indígena” é de fundamental importância, uma vez que

as unidades lexicais de uma língua se configuram como uma rica fonte de pesquisa, pois, por meio do seu estudo, detectam-se particularidades sociais, históricas e culturais de determinada comunidade linguística,

³ Considera-se a definição de ideologia defendida por Fiorin (1998, p. 28) como as “[...] representações que servem para justificar e explicar a ordem social, as condições de vida do homem e as relações que ele mantém com os outros homens [...]”. No que tange às formações discursivas, o autor destaca que “As visões de mundo não desvinculam da linguagem, porque a ideologia visto como algo imanente à realidade é indissociável da linguagem.” Dessa forma, as representações ideológicas materializam-se pelo discurso. (FIORIN, 1998, p. 33)

GUILHERME, Maurício Rubens de Carvalho; WALDOLATO, Rafaela Amaral. Um estudo da variação lexical dos itens “índio” e “indígena” em textos escritos de Língua Portuguesa. *Revista Intercâmbio*, v.LVII, 68908, 2025. São Paulo: LAEL/PUCSP. ISSN 2237-759X

pelo fato de que uma das características principais do léxico é justamente a sua dinamicidade e variabilidade, não só no tempo, mas também no espaço geográfico e nos diversos extratos sociais.(BÓZIO; BUSSE, 2014, s.d.)

Para tanto, objetivou-se quantificar a frequência de cada um dos itens lexicais em corpora compostos por textos escritos em língua portuguesa do Brasil e de Portugal; analisar a variação presente nos domínios discursivos acadêmico, ficcional e jornalístico, assim como comparar a ocorrência desses itens no século XIX e no século XX.

A fim de elucidar o problema de pesquisa, o qual questiona como os falantes de língua portuguesa se referem aos povos originários e seus descendentes do Brasil, na descrição do fenômeno foram analisados os contextos discursivos que favorecem essa variação no nível lexical, a influência de fatores extralinguísticos e a possibilidade de mudança em curso no período explorado.

Diante das pertinentes discussões sociais a respeito das diferenças significativas em atribuir ao mesmo referente os termos “Índio” e “Indígena” e do pioneirismo desse estudo na descrição dessas variantes, acredita-se que a pesquisa irá contribuir para a análise linguística e social do uso dessas palavras e impactar positivamente na utilização e disseminação do item lexical “Indígena”.

2. Referencial Teórico

Segundo Cezario e Votre (2008) a Sociolinguística, também denominada “Sociolinguística Variacionista” ou “Teoria da Variação” foi difundida com os estudos de alguns linguistas, como William Labov e William Bright, insatisfeitos com as correntes que analisavam apenas a competência linguística e os aspectos internos da língua e se estabeleceu na década de 1960 influenciando vários estudos ao longo dos anos. De encontro ao estruturalismo e ao gerativismo que consideravam a língua em uso, seus aspectos socioculturais e as variações presentes nas mais diversas manifestações linguísticas como alheias ao objeto da Linguística, a Sociolinguística surge como uma ciência social que examina, também, os fatores externos que atuam na língua.

Ressalta-se que o estudo da língua não pode ser feito de maneira autônoma, desconsiderando o contexto situacional, cultural e sócio-histórico dos falantes e que, como um código de comunicação, a língua possui um caráter essencialmente social, adaptativo e heterogêneo (CEZARIO; VOTRE, 2008).

Dessa forma, o objeto de estudo desta área da Linguística é o fenômeno da variação, a qual pode ser descrita e analisada cientificamente, como um princípio geral e universal, sistemático e estatisticamente previsível (MOLLICA, 2010). Para tanto, "Possui uma metodologia bem delimitada que fornece ao pesquisador ferramentas para estabelecer variáveis, para coleta e codificação dos dados, bem como instrumentos computacionais para definir e analisar o fenômeno variável que se quer estudar" (CEZARIO; VOTRE, 2008, p. 141-142).

Diante da diversidade e da variabilidade inerentes aos sistemas linguísticos, Cezario e Votre (2008) explicam que as variantes são identificadas pela maneira alternativa para designar um referencial com o mesmo significado básico, como "aipim", "mandioca" e "macaxeira"; "nós" e "a gente"; "os menino" e "os meninos". Como variáveis compreende-se o fenômeno em variação e o grupo de fatores atuantes nesta variação, os quais estão presentes em todos os níveis da língua - lexical, gramatical, fonético-fonológico, discursivo e semântico - e fora do sistema linguístico, como os fatores regionais, sociais e de registro - sexo, escolarização, nível de renda, grau de formalidade, dentre outros.

De acordo com Mollica (2010), o sociolinguista estuda a língua em seu uso real, natural e espontâneo, buscando descrever os fatores estruturais e sociais que motivam a alternância de uso das formas, a fim de compreender como a variação se caracteriza e o seu nível de comprometimento como fenômeno variável no sistema linguístico. Dada a importância de cada variável, identificam-se as variantes em competição e o seu grau de estabilidade ou mutabilidade ao longo do tempo - diacronismo - ou em um tempo específico - sincronismo.

Salienta-se que essas variáveis nem sempre se apresentam de forma isolada, uma vez que várias delas podem interferir no comportamento de uma mesma variação (CEZARIO; VOTRE, 2008). Portanto, a variação e a mudança ocorrem de maneira contextualizada, as quais apresentam diferentes parâmetros e diversos condicionantes que concorrem simultaneamente para o emprego de determinada forma, sendo afetados por fatores dentro e fora do sistema linguístico (MOLLICA, 2010).

Cezário e Votre (2008, p. 151) evidenciam que a análise dessas variáveis pode indicar a existência de "estabilidade entre as variantes" - variação - ou "a competição entre as variantes com o aumento de uso de uma das variantes" - mudança em curso. No entanto,

Para ocorrer uma mudança linguística [...] é necessária a interferência de fatores sociais, refletindo as lutas de poder, o prestígio entre classes,

sexos e gerações. Mas, para ocorrer a mudança, é necessário um período de variação entre formas. (CEZÁRIO; VOTRE 2008, p. 151)

Nesse sentido, os autores destacam que, com o apoio dos dados estatísticos, a análise dos fatores linguísticos e extralinguísticos observados nas diferentes comunidades de fala permite ao sociolinguista a interpretação dos aspectos que propiciam ou desfavorecem o uso de uma variante.

Quanto às variações lexicais, Tarallo (2001 apud MACHADO FILHO 2014, p. 273) considera "cada forma diferente de se representar, em um mesmo contexto, um mesmo valor significativo ou funcional, independentemente de as alterações na forma terem origem fonética, fonológica, morfológica, sintática ou discursiva." Enfatiza-se, pois, que o léxico é considerado uma riqueza sociocultural dos falantes de uma língua e evidencia suas trajetórias históricas, refletindo a relação do homem com o seu meio. Dessa forma, pode ser definido como

uma entidade abstrata que se obtém por acumulação: às palavras em uso por cada falante, no seio de uma dada comunidade de falantes, juntam-se as palavras em uso por outras comunidades linguísticas falantes da mesma língua; às palavras em uso na contemporaneidade, somam-se as que estiveram em uso em sincronias passadas, de que temos notícia pela documentação escrita e que, por vezes, ressurgem; aos dados da escrita, unem-se os dados da oralidade, quando é possível apreendê-la, dada a muito maior fluidez da oralidade face à escrita. (VILLALVA E SILVESTRE 2014, p.23 apud NETO, 2018, s.p.)

Portanto, "a língua, em especial no nível lexical, sofre influências do ambiente, que individualiza os grupos no que tange aos aspectos culturais e sociais." Assim, quando o falante utiliza unidades lexicais para nomear elementos reais, essas refletem as características de sua comunidade de fala, como a visão de mundo, questões relacionadas à etnia, a crenças e a fatores socioculturais (BÓZIO; BUSSE, 2014, s.p.).

3. Metodologia

A abordagem do problema de pesquisa foi feita de maneira quantitativa e qualitativa, uma vez que, conforme Mascarenhas (2012) há a coleta e a análise dos dados obtidos, os quais foram quantificados, garantindo a confiabilidade e a capacidade de generalização para a conclusão dos estudos. Outrossim, o autor destaca que a abordagem qualitativa auxilia na descrição do objeto analisado com mais profundidade, visando à sua compreensão. Dessa forma, ela se torna

uma ferramenta importante para os estudos sobre o comportamento dos indivíduos e de seu grupo social.

De acordo com o objetivo geral, o presente trabalho aborda a pesquisa descritiva, a qual tem como finalidade “descrever as características de uma população ou fenômeno, além de analisar se há relação entre as variáveis analisadas.” Ademais, por meio da pesquisa exploratória, há o levantamento bibliográfico baseado na teoria da variação, a qual fundamenta a descrição dos fatos. (MASCARENHAS, 2012, p. 47)

A pesquisa foi realizada por meio de dados secundários presentes no site www.corpusdoportugues.org, criado pelo professor e linguista americano Mark Davies e financiado pelo *National Endowment for the Humanities*. O site apresenta seus corpora divididos em três categorias: “Genre / Historical”, “Web / Dialects” e “NOW (2012 - 2019)”. A pesquisa foi feita na categoria “Genre / Historical”, em que se encontram corpus separados por variação dialetal do português do Brasil e de Portugal; por séculos, possibilitando uma análise diacrônica e por gêneros textuais, contribuindo para a análise da variação dos itens lexicais “Índio” e “Indígena”, nos mais diversos contextos linguísticos.

A categoria em questão separa a frequência dos dados em séculos, em que são analisados corpus do século XIX e do século XX; em dados do português de Portugal e dados do português do Brasil; nos domínios discursivos acadêmico, ficcional e jornalístico. No site também estão presentes alguns dados de oralidade, mas como objetiva-se analisar textos escritos em língua portuguesa, esses dados não foram incluídos no trabalho.

Nesse sentido, foram consideradas as variáveis dependente e independentes. A variável dependente, ou variável resposta, é aquela que se pretende analisar, isto é, aquela que sofrerá influências dos dados das outras variáveis, chamadas de independentes. Neste estudo, a variável dependente é composta pelas variantes “Índio” e “Indígena”. Ressalta-se que como a flexão de número dos itens lexicais – singular e plural – não interfere nos resultados pretendidos na pesquisa, ambos foram considerados no corpus de análise. Assim, a variável resposta também é composta pela ocorrência do plural dos termos - “Índios” e “Indígenas”.

Segundo Oliveira (2009 apud. GUILHERME, 2021), dadas apenas duas possibilidades de variável resposta, considera-se “Indígena” como a variante inovadora, a qual se pretende testar o uso entre os falantes de língua portuguesa. Por uma questão técnica do programa estatístico essa é codificada com o código 1 e recebe o nome de *sucesso*. Em

contrapartida, a variante padrão “Índio” é denominada *fracasso* e lhe é atribuído o código 0.

Já as variáveis independentes, ou variáveis explicativas, são aquelas que ocasionam a ocorrência do fenômeno que se pretende descrever. As variáveis explicativas que foram analisadas são: Domínio Discursivo, Período e Dialeto. Quanto ao domínio discursivo, estão presentes os fatores Acadêmico, Ficcional e Jornalístico. No período, foram exploradas as ocorrências das variantes no século XIX e no século XX. Em relação ao dialeto, abordou-se o fator português do Brasil e o fator português de Portugal.

Além disso, identificar a categoria na qual os itens estavam inseridos foi de fundamental importância, uma vez que esses podem ser empregados em classes gramaticais diferentes - substantivos ou adjetivos – o que afeta o resultado da análise.

Diante disso, todos os trechos dos textos que continham as palavras “Índio/Índios” e “Indígena/Indígenas” disponíveis no site foram examinados e classificados em uma planilha do Excel, a qual serviu de base para a organização dos dados necessários às funções estatísticas aplicadas na pesquisa.

Para análise dos resultados, utilizou-se o pacote estatístico da linguagem de programação denominada R, o qual, segundo Guilherme (2021), permite que a análise dos dados obtidos na pesquisa sociolinguística seja realizada de forma gratuita, estando o software disponível em diversos sistemas operacionais, como no Windows e no MacOS. O autor também destaca que existem materiais de apoio que auxiliam no entendimento dos processos, fornecendo um passo a passo da rotina de tratamento dos dados na plataforma, fundamentais para os usuários que não são familiarizados com esse tipo de linguagem.

A fim de verificar a influência de cada fator analisado sobre o fenômeno, o pacote estatístico forneceu à pesquisa o Peso Relativo, um parâmetro compreendido entre os valores 0 e 1 que permite interpretar se determinada variável explicativa favorece ou desfavorece a ocorrência da variante inovadora. Nesse sentido, quando o peso relativo de um fator for igual a 0.5 considera-se que esse fator não favorece nem desfavorece essa forma variante. Os pesos relativos inferiores a 0.5 demonstram o desfavorecimento da variante, enquanto pesos relativos maiores que 0.5 indicam o seu favorecimento (GUILHERME, 2021).

4. Análise dos resultados

Os corpora disponíveis no site www.corpusdoportugues.org forneceram à pesquisa 2.502 ocorrências dos itens lexicais em análise. Dessas ocorrências, 1.824 correspondem à variante “Índio” e 678 à variante “Indígena”. Quanto ao domínio discursivo, 723 itens estavam presentes no domínio Acadêmico, 1.326 no Ficcional e 453 no Jornalístico. Nos períodos analisados, o século XIX contou com 794 ocorrências e o século XX com 1.708 delas. De acordo com o dialeto, o português do Brasil representa 1.965 itens e o português de Portugal 537. O Quadro 1 demonstra as ocorrências acima citadas.

Quadro 1: Total das ocorrências

Variante	Domínio Discursivo	Período	Dialeto
Índio: 1.824 Indígena: 678	Acadêmico: 723 Ficcional: 1326 Jornalístico: 453	Século XIX: 794 Século XX: 1.708	Brasil: 1.965 Portugal: 537

Fonte: Elaboração autores

Observa-se, nesse sentido que, de forma geral, o uso da variante “Índio” foi significativamente maior que o uso da variante “Indígena”, representando 73% das ocorrências, enquanto este representa 27% delas. É importante salientar, contudo, que nem sempre os itens analisados correspondem à mesma classe gramatical, podendo ser empregados como núcleo do sintagma nominal – substantivo – ou como modificador – adjetivo. Retomando o problema de pesquisa do trabalho – como os falantes de língua portuguesa se referem aos povos originários e seus descendentes do Brasil? – a categorização dos itens lexicais se torna fundamental para a análise dos resultados.

Diante disso, dos termos contabilizados, 1.997 eram substantivos e 505 eram adjetivos. Ressalta-se que desses 1.997 substantivos, apenas 11% correspondem à forma “Indígena” e 89% correspondem a “Índio”. No que tange aos 505 adjetivos, 90,4% das ocorrências eram de “Indígena” e 9,6% eram de “Índio”. Pode-se observar a representação desses dados na Tabela 1.

Tabela 1: As categorias Substantivo e Adjetivo

Categoria	Total	%	PR
Substantivo	1.997	11%	0.1
Adjetivo	505	90,4%	0.9

Fonte: Elaboração autores

A Tabela 1 representa os dados da categoria – substantivo e adjetivo – referentes à variante que se pretende investigar, ou seja, o item lexical “Indígena”. Quanto ao peso relativo, constata-se que ambos os fatores apresentam valores diferentes de 0.5, portanto, ser adjetivo favorece a ocorrência de “Indígena” com $PR=0.9$ e ser substantivo desfavorece a ocorrência desse item, com $PR=0.1$. Esse resultado, juntamente com o número total de ocorrências das variantes, confirma uma das hipóteses levantadas no trabalho, de que o uso do item lexical “Índio” para se referir aos povos originários e seus descendentes do Brasil é mais frequente na língua portuguesa.

Diante do exposto, julgou-se necessário analisar mais detalhadamente as ocorrências desses itens em posição de núcleo do sintagma nominal, ou seja, substantivos, já que essa é a estrutura usada pelos falantes para nomear os povos originários e seus descendentes.

Em relação aos substantivos contabilizados, do total de 1.997 ocorrências, 1.774 itens correspondem à variante “Índio” e 223 à variante “Indígena”. Desse total, 483 estão presentes no domínio discursivo Acadêmico, 1.177 no Ficcional e 337 no Jornalístico. O século XIX representa 694 das variantes verificadas e o século XX, 1.303. Em relação ao Dialeto, o português do Brasil corresponde a 1.615 ocorrências e o português de Portugal a 382 delas, conforme mostra o Quadro 2.

Quadro 2: Total dos resultados com substantivos

Variante	Domínio Discursivo	Período	Dialeto
Índio: 1.774 Indígena: 223	Acadêmico: 483 Ficcional: 1.177 Jornalístico: 337	Século XIX: 694 Século XX: 1.303	Brasil: 1.615 Portugal: 382

Fonte: Elaboração autores

Percebe-se que a variável resposta apresenta valores expressivamente discrepantes, em que a variante padrão corresponde a 89% das ocorrências. Portanto, é imprescindível a descrição dos fatores que favorecem ou desfavorecem o uso da variante inovadora. A tabela a seguir evidencia os dados da variável Domínio Discursivo com relação à variável dependente.

Tabela 2: Variável Domínio Discursivo

Domínio Discursivo	Total	%	PR
Acadêmico	483	18,8	0.64
Ficcional	1177	9,9	0.50
Jornalístico	337	4,5	0.33

Fonte: Elaboração autores

No domínio discursivo Acadêmico, 18,8% dos itens contabilizados corresponde à variante “Indígena” e 81,2% à variante “Índio”. Já o fator Ficcional, em que se observa a maior frequência de dados analisados, apenas 9,9% deles representa a variável *indígena* e 90,1% representam a variável *índio*. O fator Jornalístico apresenta um total de 337 itens, dos quais 4,5% equivalem à forma variante “Indígena” e 95,5% à forma variante “Índio”.

Quanto ao peso relativo e ao favorecimento ou desfavorecimento desses fatores para a ocorrência da variante “Indígena”, nota-se que o domínio discursivo Acadêmico a favorece, com $PR=0.64$; o Jornalístico a desfavorece com $PR=0.33$ e o fator Ficcional não favorece nem desfavorece tal ocorrência, $PR=0.5$.

O favorecimento da variante inovadora pelo domínio discursivo Acadêmico corrobora outra hipótese levantada no início da pesquisa de que em ambientes linguísticos mais formais a escolha pelo uso do item “Indígena” seria mais frequente, uma vez que se acreditava que as discussões sobre o valor social desses usos estaria mais presente nesses contextos.

Acrescenta-se que o nível de formalidade dos domínios discursivos pode estar correlacionado à gradação da abstração neles presente. Nota-se que a esfera Jornalística está mais atrelada ao presente, apresenta elementos reais como pessoa, tempo e espaço, que muitas vezes são reconhecidos por quem tem acesso ao texto. Tais elementos podem aproximar o leitor e garantir a esse domínio discursivo um aspecto mais concreto e informal, quando relacionado às outras esferas analisadas.

Por outro lado, por mais que a esfera Ficcional também possa apresentar os elementos pessoa, tempo e espaço e esses geralmente apresentem um caráter de verossimilhança, são apenas representações do real. Esse domínio discursivo tende a ser, então, menos concreto que o Jornalístico.

Já a esfera Acadêmica é regularmente mais abstrata, pois trata da produção e divulgação de conhecimentos, estudos e análises; a linguagem utilizada normalmente é impessoal, a qual distancia o observador do leitor, o que pode levar a um maior grau de formalidade.

Além desses fatores, a análise da variável Período e da variável Dialeto também é relevante para a pesquisa. Conforme a Tabela 3 constata-se que a mudança de tempo, do século XIX para o século XX, não favoreceu nem desfavoreceu o uso de “Indígena”.

Tabela 3: Variável Período

Período	N/Total	%	PR
Século XIX	694	9,4	0.5
Século XX	1303	12,1	0.5

Fonte: Elaboração autores

Nesse sentido, mesmo com o aumento das ocorrências totais e da porcentagem de verificações do item “Indígena” de 9,4% no século XIX para 12,1% no século XX, tal resultado não é significativo para indicar uma possível mudança em curso. Espera-se, pois, que os dados relacionados ao século XXI, ainda não disponíveis no corpus de pesquisa, assim como as demais variáveis explicativas, favoreçam o uso da variante “Indígena” – $PR > 0.5$.

Quanto à variável Dialeto, a Tabela 4 apresenta o total de substantivos presentes nos textos escritos do português do Brasil e do português de Portugal, a porcentagem de ocorrências da variante inovadora e o seu respectivo peso relativo.

Tabela 4: Variável Dialeto

Dialeto	Total	%	PR
Brasil	1.615	8,9	0.42
Portugal	382	20,7	0.58

Fonte: Elaboração autores

Como resultado da análise desse fator, tem-se que do total de 1.615 ocorrências desses itens lexicais no português do Brasil, 8,9% correspondem à forma “Indígena” e 91,1% a “Índio”. No que tange ao português de Portugal, dos 382 itens verificados, 20,7% deles se referem ao item “Indígena” e 79,3% ao item “Índio”. Quanto ao dado estatístico Peso Relativo, nota-se que no dialeto europeu há um favorecimento da ocorrência do substantivo “Indígena” com $PR = 0.58$. Em contrapartida, observa-se o seu desfavorecimento no dialeto brasileiro, com $PR = 0.42$.

Esse resultado contrariou as expectativas de que esse item seria mais usado no português do Brasil, tendo em vista que os indígenas são brasileiros. Esperava-se, pois, uma maior conscientização acerca das discussões sociais sobre o uso dessas formas variantes no país, dada a

imposição do nome “Índio” aos povos originários no processo de colonização portuguesa.

Entretanto, é possível que o favorecimento da variante inovadora no dialeto europeu possa ser explicado pela discrepante posição desses países no ranking do Índice de Desenvolvimento Humano (IDH), que tem como um dos parâmetros o nível educacional da sociedade. Nesse índice, o acesso à educação é mensurado pela média dos anos de escolaridade de pessoas adultas e pela expectativa de vida escolar das crianças quando em idade para iniciar seus estudos (PNUD, s.d.). Assim, de acordo com a Organização das Nações Unidas (ONU, 2020), “dentre as nações de língua portuguesa, Portugal é o melhor colocado no IDH com a posição 38 de 189.” O Brasil ocupa a posição 84 desse indicador.

O desenvolvimento no âmbito educacional de Portugal pode ser uma condição que propicia o uso da forma variante “Indígena”, já que, em geral, questões de natureza sociopolíticas relativas a direitos humanos e à valorização de comunidades marginalizadas são discutidas, primeiramente na academia.

Outro ponto que porventura pode justificar o favorecimento da variante “Indígena” no dialeto português de Portugal é a relação do favorecimento dessa variável - $PR=0.58$ - com o favorecimento do fator Acadêmico - $PR=0.64$ - na variável domínio discursivo. Nesse sentido, dos 382 itens que correspondem ao português de Portugal, 62,3% são do domínio discursivo Acadêmico. No dialeto brasileiro, dos 1.615 itens contabilizados, apenas 15,2% estão relacionados a essa esfera discursiva. As tabelas 5 e 6 apresentam os valores absolutos e as porcentagens de cada domínio discursivo analisado de acordo com o dialeto.

Tabela 5: Português do Brasil: domínios discursivos

Domínio Discursivo	Total	%
Acadêmico	245	15,2
Ficcional	1.055	65,3
Jornalístico	315	19,5

Fonte: Elaboração autores

Tabela 6: Português de Portugal: domínios discursivos

Domínio Discursivo	Total	%
Acadêmico	238	62,3
Ficcional	122	31,9
Jornalístico	22	5,8

Fonte: Elaboração autores

Diante disso, considera-se a possibilidade do favorecimento do item lexical “Indígena” no dialeto europeu ser uma falsa impressão. Analisando a representatividade das esferas discursivas dentro dos diferentes dialetos, observa-se que no português de Portugal a frequência de dados do domínio discursivo Acadêmico é muito mais relevante, o que pode estar interferindo no favorecimento da variante inovadora.

5. Discussões sobre os aspectos sociais do fenômeno

Segundo Dorrico (2021), cerca de seis milhões de indígenas habitavam o território hoje chamado de Brasil antes do violento processo de colonização, o qual dizimou, explorou e desconsiderou as tradições, os saberes e a cultura dos povos originários, atribuindo-lhes o nome “Índio”. Desde então, a invisibilidade, o preconceito e a desumanidade estiveram presentes na vida dessas populações ancestrais no país.

Munduruku (2018) destaca que as populações indígenas só ficaram visíveis a partir da década de 1970 e reconhecidamente visíveis em 1988 com a Constituição Federal, que em concordância com Dorrico (2021) foi um marco importante na afirmação dos direitos civis e da identidade indígena. Nesse período, inicia-se a organização do movimento indígena brasileiro, em que se destacam as lutas sociais por direitos, atuando na “formação de lideranças próprias, articulação entre os povos e parceria com entidades de apoio e com o Estado.” (UNB, 2011).

Quanto ao uso da nomenclatura “Índio” pelos povos originários, Daniel Munduruku expõe que

Nos anos 70, quando essa juventude começou a olhar, a se perceber parte de uma sociedade maior, que foi assim que começou o movimento indígena, essa juventude usou esse termo índio como uma forma de luta, como uma forma de identificação daqueles que eram parceiros. Então, essa palavra ainda é usada, e se é usada por uma liderança dessa, é nesse sentido (MUNDURUKU, 2018).

É importante reconhecer que o item lexical “Índio” está presente no vocabulário de língua portuguesa – inclusive dos povos indígenas – dada a repetição exaustiva do termo imposto. Porém, quando um indígena se utiliza do termo, ele o faz politicamente, reforçando sua identidade. Em contrapartida, a sociedade brasileira o faz como uma forma de desdém, numa categorização de menos humanidade (MUNDURUKU, 2018).

Segundo Munduruku (2019), “Índio” é uma palavra ideologizada, pejorativa e folclórica que perdeu o seu sentido. “É uma palavra que só desqualifica, remonta a preconceitos. É uma palavra genérica. Esse generalismo esconde toda a diversidade, riqueza, humanidade dos povos indígenas.” Para ele, a sociedade deveria abandonar o uso dessa palavra e utilizar a palavra “Indígena”, a qual diz muito mais a respeito dos povos originários e de seus descendentes.

No entanto, a maioria das pessoas acredita que as palavras “Índio” e “Indígena” derivam uma da outra, mas essas possuem significados diferentes e isso pode ser comprovado. De acordo com o Dicionário Aurélio (1988, p. 358) a forma “Índio” apresenta um caráter homônimo, em que se verificam dois significados distintos. O primeiro deles expõe que esse item lexical é um elemento químico da tabela periódica, apresentando suas características. A segunda ocorrência define o termo como “De, ou pertencente ou relativo ao índio” e “O habitante das terras americanas ao chegarem os descobridores europeus; o aborígine da América” (MUNDURUKU, 2018).

Sob outro enfoque, o Dicionário Aurélio (1988, p. 358) designa o item lexical “Indígena” como “Originário de um país ou de uma localidade” e “Pessoa natural do lugar ou país onde habita”. Diante do exposto, Munduruku (2018) afirma que os povos originários do Brasil são indígenas e essa denominação não se trata do que é politicamente correto, mas sim do que é efetivo. A escolha dos itens lexicais em análise pode enobrecer ou depreciar esses povos ancestrais que tanto valorizam a cultura da palavra. Portanto, utilizá-la para tratá-los como de fato são e como querem ser tratados é sinal de humanidade, de inteligência e de tolerância.

Diante do exposto, evidencia-se a necessidade de ampliação das discussões a respeito dos aspectos sociais do fenômeno analisado, o que se espera culminar na utilização do item lexical “Indígena” para se referir aos povos originários e seus descendentes do Brasil.

Nesse sentido, as redes sociais têm se tornado um importante meio para expandir as possibilidades de divulgação, compartilhamento de informações e conhecimento crítico acerca do uso da palavra “Índio” e de tantas outras formas de apagamento da história, memória e cultura dos povos indígenas.

Ademais, a produção literária indígena como símbolo de cultura e de resistência, expõe a importância dos povos originários terem sua identidade e diversidade respeitadas pela sociedade brasileira. Um exemplo significativo disso é o poema *Índio eu não sou* da escritora indígena Márcia Wayna Kambeba.

Outro fato relevante é o projeto de lei 5.466/19 da deputada federal e coordenadora da Frente Parlamentar Mista em Defesa dos Direitos dos Povos Indígenas, Joenia Wapichana. O projeto, que aguarda deliberação do senado, objetiva substituir a nomenclatura "Dia do Índio", comemorado no dia 19 de abril para "Dia dos Povos Indígenas".

Segundo Wapichana (2021), tal mudança torna a expressão "mais respeitosa e mais identificada com as comunidades indígenas". Na justificativa do projeto, a deputada considera que os povos indígenas contribuem para a "diversidade cultural, a harmonia social e ecológica da humanidade" e é "importante frisar que a contribuição é ofertada pela coletividade e não pelo indivíduo isolado, como remete a ideia do termo 'índio'".

6. Considerações Finais

Este trabalho teve como objetivo geral descrever a variação linguística entre os itens lexicais "Índio" e "Indígena" em textos escritos de língua portuguesa. A hipótese considerada ao longo da pesquisa era de que o uso do item lexical "Índio" é mais frequente na língua portuguesa, mas que se observa uma tendência crescente de substituição para a palavra "Indígena", principalmente em domínios discursivos mais formais, devido às discussões sobre o valor social desses usos.

Dentre as conclusões mais relevantes está a predominância da variante padrão na amostra analisada. De um total de 2.502 itens lexicais, 73% desses correspondiam à forma "Índio" e 23% à forma "Indígena". Quanto ao uso da variável resposta como substantivo – objeto da pesquisa – 89% das ocorrências eram representadas pela variante "Índio" e apenas 11% pela variante inovadora. Por outro lado, observou-se na categoria adjetivo que "Indígena" abarcou 90,4% das ocorrências. Nesse sentido, ser adjetivo favoreceu a ocorrência de "Indígena", com PR=0.9 e ser substantivo desfavoreceu essa variante, com PR=0.1.

Quanto à esfera discursiva, o fator Acadêmico favoreceu o uso da variante *indígena*, com PR= 0.64 e o fator Jornalístico a desfavoreceu, com PR=0.33. Em relação à variável Período, nos séculos XIX e XX não houve o favorecimento nem o desfavorecimento da forma inovadora. Diante desses dados, depreende-se que o item lexical "Índio" é mais frequente na língua portuguesa e que mesmo com o favorecimento do item lexical "Indígena" na esfera discursiva Acadêmica, não há uma tendência de mudança em curso no período analisado.

Outro resultado importante é o favorecimento da ocorrência da forma “Indígena” no dialeto europeu, com PR=0,58 e o seu desfavorecimento no português do Brasil, com PR=0.42. No entanto, considera-se a possibilidade de tal fato ser uma falsa impressão quando se correlaciona a frequência de dados do domínio discursivo Acadêmico com o dialeto em questão.

A respeito do uso dos itens lexicais “Índio” e “Indígena” para se referir aos povos originários e seus descendentes do Brasil, acredita-se que a pesquisa irá contribuir para a análise linguística e social desse fenômeno, impactando positivamente na utilização do substantivo “Indígena”.

Como a pesquisa não apresenta dados do século XXI, é esperado que as discussões e conclusões desse artigo sejam relevantes para trabalhos futuros, ampliando as possibilidades de análise dessas variantes. Dessa forma, evidencia-se que a riqueza e a diversidade vocabular presentes na língua portuguesa e as influências externas que atuam na variação em nível lexical a constituem como um importante campo de pesquisa da Sociolinguística.

Referências bibliográficas

BÓZIO, Jéssyca Finantes do Carmo; BUSSE, Sanimar. Variação lexical: um olhar para a formação cultural do falante. In: *Revista Línguas e Letras*. vol. 15, n. 31, 2014. Disponível em: <<https://e-revista.unioeste.br/index.php/linguaseletras/article/view/11537/8265>> Acesso em: 13 nov. 2021.

CEZARIO, Maria Maura; VOTRE, Sebastião. Sociolinguística. In: MARTELOTTA, Mário Eduardo (Org.). *Manual de Linguística*. São Paulo: Contexto, 2008. Disponível em: <<https://plataforma.bvirtual.com.br/Leitor/Publicacao/1390>> Acesso em: 14 nov. 2021.

DAVIES, Mark. *O CORPUS do português*. 2021. Disponível em: <www.corpusdoportugues.org> Acesso em: 12 dez. 2021

DORRICO, Julie. *Literatura e Ancestralidade*. Junho de 2021. 103 minutos. Plataforma Youtube. Disponível em: <<https://youtube.com/watch?v=y1-gzFJOtNI&feature=share>> Acesso em: 20 mai. 2022.

FERREIRA, Aurélio Buarque de Holanda. *Dicionário Aurélio Básico de Língua Portuguesa*. Rio de Janeiro: Nova Fronteira, 1988.

GUILHERME, Maurício Rubens de Carvalho; WALDOLATO, Rafaela Amaral. Um estudo da variação lexical dos itens "índio" e "indígena" em textos escritos de Língua Portuguesa. *Revista Intercâmbio*, v.LVII, 68908, 2025. São Paulo: LAEL/PUCSP. ISSN 2237-759X

FERREIRA, Lenne; LACERDA, Victor. Resistência Indígena: Entenda porquê o termo "índio" é considerado pejorativo. In: *Alma Preta – Jornalismo Preto e Livre*. 2021. Disponível em: <<https://almapreta.com/sessao/cotidiano/resistencia-indigena-entenda-porque-o-termo-indio-e-considerado-pejorativo>> Acesso em: 13 nov. 2021.

FIORIN, José Luiz. *Linguagem e Ideologia*. 6 ed. São Paulo: Ática. 1998. Disponível em: <https://www.academia.edu/37986651/Linguagem_e_Ideologia_Jose_Luiz_Fiorin> Acesso em: 17 jun. 2022.

GUILHERME, Maurício Rubens de Carvalho. *Lexicalização-Degramaticalização do plural do português falado em Belo Horizonte*. 2021. Tese de doutorado. Faculdade de Letras, Universidade Federal de Minas Gerais. Belo Horizonte, 8 Fev. 2021. Disponível em: <<https://repositorio.ufmg.br/handle/1843/35701>> Acesso em: 12 dez. 2021.

KAMBEBA, Márcia. Índio eu não sou. In: *DORRICO, Julie. 3 poemas de Márcia Kambeba*. 2020. Disponível em: <<https://revistaacrobata.com.br/julie-dorrico/poesia/3-poemas-de-marcia-kambeba/>> Acesso em: 01 jul. 2022.

MACHADO FILHO, Américo Venâncio Lopes. Do conceito de "variante" nos estudos do léxico de perspectiva histórico-variacional. In: *Portal de Revistas da USP. Filologia e linguística portuguesa*. 2014. Disponível em: <<https://www.revistas.usp.br/flp/article/view/83852>> Acesso em: 4 dez. 2021.

MASCARENHAS, Sidnei Augusto. *Metodologia científica*. Pearson EducationdoBrasil, 2012. p. 45-47.

MOLLICA, Maria Cecília. Fundamentação teórica: conceituação e delimitação. In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza. (Org.) *Introdução à sociolinguística: o tratamento da variação*. 4 ed. São Paulo: Contexto, 2010.

MUNDURUKU, Daniel. Índio ou Indígena. 2018. In: *ITAÚ Cultural. Mekukradjá Círculo de Saberes*. Plataforma Youtube. 5:32 min. Disponível em: <<https://www.youtube.com/watch?v=s39FxY3JzIE>> Acesso em: 10 jun. 2022.

GUILHERME, Maurício Rubens de Carvalho; WALDOLATO, Rafaela Amaral. Um estudo da variação lexical dos itens "índio" e "indígena" em textos escritos de Língua Portuguesa. *Revista Intercâmbio*, v.LVII, 68908, 2025. São Paulo: LAEL/PUCSP. ISSN 2237-759X

MUNDURUKU, Daniel. Dia do índio é data "folclórica e preconceituosa", 2019. In: *G1 Educação*. Disponível em: <<https://g1.globo.com/educacao/noticia/2019/04/19/dia-do-indio-e-data-folclorica-e-preconceituosa-diz-escriptor-indigena-daniel-munduruku.ghtml>> Acesso em: 21 nov. 2021.

NETO, João Irineu de França. Variação lexical do português: um estudo comparativo entre o Atlas Linguístico-etnográfico de Portugal e da Galiza e atlas regionais brasileiros. 2018. Disponível em: <<https://www.revistas.uneb.br/index.php/comsertoes/article/view/5900>> Acesso em 26 nov. 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Índice de Desenvolvimento Humano da ONU inclui variante pegada de carbono. 2020. Disponível em: <<https://news.un.org/pt/story/2020/12/1736222>> Acesso em: 28 jun. 2020.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD). O que é IDH. Disponível em: <<https://www.undp.org/pt/brazil/o-que-%C3%A9-o-idh>> Acesso em: 04 de jul. 2022.

UNIVERSIDADE DE BRASÍLIA. Pesquisa recupera história do movimento indígena no Brasil. 2011. Disponível em: <<https://unbciencia.unb.br/humanidades/94-historia/280-pesquisa-recupera-historia-do-movimento-indigena-no-brasil>> Acesso em: 15 jun. 2022.

WAPICHANA, Joenia. Projeto de Lei 5466/19 da deputada Joenia Wapichana que institui Dia dos Povos Indígenas é aprovado na CCJC da Câmara. 2021. Disponível em: <<https://www.joeniawapichana.com.br/imprensa/projeto-lei-5466-19-da-deputada-joenia-wapichana-que-institui-dia-dos-povos-indigenas-e-aprovado-na-ccjc-da-camara>> Acesso em: 15 jun. 2022.

Recebido: 29/10/2024
Aprovado: 06/10/2025



Esta obra está licenciada com uma Licença [Creative Commons Atribuição 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/) que permite o uso irrestrito, distribuição e reprodução em qualquer meio, desde que a obra original seja devidamente citada