

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP**

LUIZ CARLOS ZEFERINO

**REPRESENTAÇÕES VISUAIS DO BRASIL:
UMA ABORDAGEM MULTIDIMENSIONAL IMAGÉTICA**

MESTRADO EM LINGUÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

**SÃO PAULO
2023**

LUIZ CARLOS ZEFERINO

**REPRESENTAÇÕES VISUAIS DO BRASIL:
UMA ABORDAGEM MULTIDIMENSIONAL IMAGÉTICA**

MESTRADO EM LINGUÍSTICA APLICADA E ESTUDOS DA LINGUAGEM

Dissertação apresentada à Banca Examinadora da Pontifícia Universidade Católica de São Paulo, como exigência parcial para obtenção do título de Mestre em Linguística Aplicada e Estudos da Linguagem sob orientação do Professor Doutor Antonio Paulo Berber Sardinha.

São Paulo

2023

BANCA EXAMINADORA

Prof. Dr. Tony Berber Sardinha

Prof. Dr. Carlos Henrique Kaufmann

Prof. Dr. Cristiano Burmester

Profa. Dra. Elizabeth Brait

Profa. Dra. Maria Cláudia Delfino

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação de mestrado, por processos fotocopiadores ou eletrônicos, desde que devidamente citada.

Assinatura:

A handwritten signature in black ink, appearing to read 'Luiz Carlos Zeferino'. The signature is fluid and cursive, with a horizontal line extending from the left side.

Data: 30 de julho de 2023

FICHA CATALOGRÁFICA

ZEFERINO, Luiz Carlos. São Paulo, 2023, pp. 123.

Dissertação de Mestrado – Pontifícia Universidade Católica de São Paulo.

Área de Concentração: Linguística Aplicada e Estudos de Linguagem.

Orientador: Professor Doutor Antonio Paulo Berber Sardinha.

REPRESENTAÇÕES VISUAIS DO BRASIL: UMA ABORDAGEM MULTIDIMENSIONAL IMAGÉTICA.

Luiz Carlos Zeferino

RESUMO: "Brasilidade" é a palavra para a qual deve-se o tema e a ideia deste trabalho de pesquisa. Servindo quase como um rótulo, o termo serve de adjetivo ou sinônimo àqueles atributos os quais traduzem anos de alteridade e da impressão indelével, daquilo que identifica a imagem brasileira. Esta pesquisa nasce, também, de acordo com as novas práticas e os pilares da Ciência Aberta (BERBER SARDINHA et al., 2021), numa análise discursiva e multidimensional (BIBER, 1988), em Linguística de Corpus (BERBER SARDINHA, 2004), a respeito da construção do imaginário discursivo e histórico dessa "brasilidade". Propor essa reflexão no campo da comunicação visual envolve mapear os aspectos multidimensionais dessa problemática, analisando o corpus que detém uma livre circulação, por estar na internet, de livre acesso no site-portfolio Flickr. O objetivo é a exposição de um estudo que mapeia a construção de brasilidade para a imagem brasileira, enquanto rótulo-mercadoria, ou discurso-produto. O termo não apenas imprime uma identidade, contendo atributos característicos da gestão de branding (HULBERG, 2006), mas também mobiliza as disposições e percepções a respeito do país, de acordo com a estética vigente em cada época. Metodologicamente, o estudo baseia-se na Linguística de Corpus (BERBER SARDINHA, 2004) ao analisar as imagens do corpus estudado, que deu como resultado 09 dimensões, ou seja, 09 maneiras diferentes de representação do Brasil. Essa pesquisa possibilita ainda o incentivo à sua replicação investigativa e melhorias, por meio do livre acesso aos dados abertos, seus corpora e códigos, no Portal Multimodal/Multilíngue para o avanço da Ciência Aberta (BERBER SARDINHA et al, 2021; CiênciaAberta.Org).

Palavras-Chave: Linguística de Corpus; Multimodalidade; Análise Multidimensional; Estudos da Imagem e da Cultura.

BRAZIL'S VISUAL REPRESENTATIONS: AN IMAGETIC MULTIDIMENSIONAL APPROACH

Luiz Carlos Zeferino

ABSTRACT: "Brasilidade" is the word which represents the main idea of this research. It almost works as a label, an adjective or synonym to attributes which can be seen as years of difference and unbreakable expression, identifying the Brazilian image. This research was also born according to new practice and the main objectives of Open Science (BERBER SARDINHA et al., 2021), on a discursive and multimodal (BIBER, 1988), in Corpus Linguistics (BERBER SARDINHA, 2004), around the construction of the discursive and historical image of this "brasilidade". To propose this reflection on the field of visual communication involves mapping the multidimensional aspects of this corpus, analyzing a corpus which is freely available because it is on the internet, on the website Flickr. The objective is the exposition of a study which maps the building of the term brasilidade for the Brazilian image, while label-merchandise, or discourse-product. The word not only shows an identity, having attributes which are characteristics of the branding world (HULBERG, 2006), but also puts into motion perceptions about the country, according to the aesthetics present in each time. Methodologically, the study is based on Corpus Linguistics (BERBER SARDINHA, 2004) when the images of the corpus are analyzed, unveiling 09 dimensions, i.e., 09 different ways to represent Brazil. This research aims to give support to investigative replication and improvement of science, through free access to data, corpus and codes, on the Portal Multimodal/Multilíngue para o avanço da Ciência Aberta (BERBER SARDINHA et al, 2021; CiênciaAberta.Org).

Keywords: Corpus Linguistics; Multimodality; Multidimensional Analysis; Image and Culture Studies.

O presente trabalho foi realizado com apoio da Conselho Nacional de Pesquisa - Brasil (CNPq) – Código de Financiamento / Chamada do Edital 25/2020.

AGRADECIMENTOS

Expresso os agradecimentos ao meu orientador, Professor Dr. Tony Berber Sardinha, de quem recebi dois anos de orientação, pois sem essa cooperação este trabalho não teria existido. À PUC-SP e Ao CNPq, pelo financiamento de minha pesquisa, como parte integrante do projeto do Portal Multimodal para o desenvolvimento da Ciência Aberta nas Humanidades Digitais.

E agradeço imensamente a cada membro do GELC, mais especialmente na pessoa do Carlos Kauffmann, Cláudia Delfino e Carol Andrade. Devo este trabalho a vocês.

Com muita alegria, dedico este volume à Maria Aparecida Zeferino, Romeu Pereira Zeferino, Luiza Pereira Zeferino e para Bianca Baroni.

LISTA DE ABREVIações

ANOVA = *Analysis of Variance*, ou Análise de Variância

br = *Brazil*, Brasil

de = Deutschland. Germany, Alemanha.

gb = Grã-Bretanha

GELC = Grupo de Estudos de Linguística de Corpus, da PUC-SP

GLM = *General Linear Model*, Modelo Linear Geral, na matemática e estatística.

LC = Linguística de Corpus

AMD = Análise Multidimensional

VPN = Rede Privada Virtual, do inglês *Virtual Private Network*

JSON = *JavaScript Object Notation*, um formato compacto de troca de dados

IA = Inteligência Artificial; AI

SAS = *Software Statistical Analysis System*, de análises de dados

p = Valor-p, a base comprobatória

pp = Páginas

in prep = No prelo; em preparação.

R = R^2 ou R Quadrado. Representa a porcentagem de variação na resposta

us = United States, USA ou EUA.

LISTA DE FIGURAS

Figura 1 - Incidência média da dimensão no Google Books Corpus.....	26
Figura 2 - Output do Google Visual Cloud AI para uma imagem do Cristo Redentor.	35
Figura 3 - Scree plot dos eigenvalues da análise fatorial.....	36
Figura 4 - Representação típica do Brasil na dimensão 1, polo negativo.....	47
Figura 5 - Representação do Brasil na dimensão 1, polo negativo.....	48
Figura 6 - Representação típica do Brasil na dimensão 1, polo negativo.....	48
Figura 7 - Representação típica belicosa (Grã-Bretanha) na dimensão 1, polo positivo ...	49
Figura 8 - Representação do Brasil no Fator 1, polo positivo	50
Figura 9 - Representação do Brasil no Fator 1, polo positivo	51
Figura 10 - Representação típica do Brasil na dimensão 2, polo positivo	56
Figura 11 - Representação típica do Brasil na dimensão 2, polo positivo	57
Figura 12 - Representação típica do Brasil na dimensão 2, polo positivo	57
Figura 13 - Representação de imagem típica da dimensão 2, polo negativo	58
Figura 14 - Representação de imagem típica da dimensão 2, polo negativo	58
Figura 15 - Representação de imagem típica da dimensão 2, polo negativo	59
Figura 16 - Representação típica do Brasil na dimensão 3, polo positivo	62
Figura 17 - Representação de Brasil no fator 3, polo positivo.....	62
Figura 18 - Representação típica do Brasil na dimensão 3, polo positivo	63
Figura 19 - Representação típica do Brasil na dimensão 3, polo positivo	63
Figura 20 - Representação de Brasil, dimensão 3, polo negativo.....	65
Figura 21 - Representação típica do Brasil na dimensão 3, polo negativo.....	65
Figura 22 - Representação típica do Brasil na dimensão 3, polo negativo.....	66
Figura 23 - Cena representando o Brasil no polo positivo do fator 4.....	70
Figura 24 - Cena urbana representando o Brasil. Fator 4, polo positivo.	70
Figura 25 - Imagem de representação do Brasil, no discurso do polo negativo, fator 4 ...	71
Figura 26 - Imagem de animais representando o Brasil no polo negativo do quarto fator. 71	
Figura 27 - Imagem representando o Brasil no polo negativo do quarto fator.....	72
Figura 28 - Representação de Brasil no polo negativo do fator 4.....	72
Figura 29 - Representação de Brasil no polo negativo do fator 4.....	73
Figura 30 - Representação de Brasil no fator 5.....	77
Figura 31 - Representação de Brasil no Fator 5.....	78
Figura 32 - Imagem representando o Brasil no polo positivo do fator 6.....	81
Figura 33 - Representação de Brasil, polo positivo do fator 6.....	81
Figura 34 - Representação de Brasil, polo negativo do fator 6.....	82
Figura 35 - Representação de imagem típica do Brasil na dimensão 6, polo negativo.....	82
Figura 36 - Imagem de representação do Brasil, polo negativo do fator 6.....	83
Figura 37 - Imagem de Brasil no polo positivo do fator 7.....	86
Figura 38 - Imagem de Brasil no polo positivo do fator 7.....	87
Figura 39 - Representação do Brasil no polo positivo do fator 7.....	87
Figura 40 - Representação de Brasil no polo negativo do fator 7.....	88
Figura 41 - Representação de Brasil no polo negativo do fator 7.....	89
Figura 42 - Representação de Brasil no polo negativo do fator 7.....	89
Figura 43 - Discurso presente no fator 8, polo positivo.	91
Figura 44 - Discurso presente no fator 8, polo positivo	92
Figura 45 - Representação de Brasil, Fator 8 - polo positivo	92
Figura 46 - Representação de Brasil, Fator 8 - polo positivo	93

Figura 47 - Representação de Brasil, Fator 8, polo positivo.	93
Figura 48 - Representação de Brasil, polo negativo do fator 8	94
Figura 49 - Representação de Brasil no polo negativo do oitavo fator	95
Figura 50 - Representação de Brasil no polo negativo do fator 8	95
Figura 51 - Representações de Brasil, polo negativo no Fator 8	96
Figura 52 - Representação de Brasil, polo positivo do fator 9.....	98
Figura 53 - Representação de Brasil, polo positivo do fator 9.....	99
Figura 54 - Representação de Brasil, polo positivo do fator 9.....	99
Figura 55 - Representação de Brasil, polo positivo do fator 9.....	100
Figura 56 - Representação de Brasil, polo positivo do fator 9.....	101
Figura 57 - Representação de Brasil, polo positivo do fator 9.....	101
Figura 58 - Representação de Brasil, polo negativo do fator 9	102
Figura 59 - Representação de Brasil, polo negativo do fator 9	102
Figura 60 - Representação de Brasil, polo negativo do fator 9	103
Figura 61 - Representação de Brasil, polo negativo do fator 9	103
Figura 62 - Representação de Brasil, polo negativo do fator 9	104
Figura 63 - Imagem do polo negativo, no corpus da Alemanha, do fator 9	105
Figura 64 - Gráfico piramidal com a variação e difusão discursiva no corpus Flickr	106
Figura 65 - Corpus Flickr, com 5833 itens JPG	122

LISTA DE TABELAS

Tabela 1 - Tipologia do Corpus analisado nesta pesquisa	33
Tabela 2 - Síntese de conceitos envolvidos na análise fatorial e multidimensional	37
Tabela 3 - Títulos curtos de cada variável discursiva	40
Tabela 4 - Polo positivo do Fator 1	41
Tabela 5 - Polo negativo do Fator 1	44
Tabela 6 - Escores médios de cada país no fator 1	46
Tabela 7 - Polo positivo do Fator 2	51
Tabela 8 - Polo negativo do Fator 2	54
Tabela 9 - Escores médios de cada país no fator 2	56
Tabela 10 - Variação do discurso no polo positivo do fator 3	59
Tabela 11 - Fator 3, carga fatorial do polo negativo	64
Tabela 12 - Fator 4 e suas variáveis do polo positivo	66
Tabela 13 - Polo negativo do quarto fator	73
Tabela 14 - Escores médios de cada país no fator 4	74
Tabela 15 - Fator 5, com um único polo	74
Tabela 16 - Escores de cada país no quinto fator	79
Tabela 17 - Fator 6, variáveis do polo positivo	79
Tabela 18 - Polo negativo do fator 6	83
Tabela 19 - Escores da dimensão 6, por país	84
Tabela 20 - Variação discursiva no polo positivo do Fator 7	84
Tabela 21 - Polo negativo do fator 7	88
Tabela 22 - Escores médios de cada país no fator 7	90
Tabela 23 - Discursos presentes no polo positivo do fator 8	90
Tabela 24 - Discursos presentes no oitavo fator negativo	93
Tabela 25 - Escores médios de cada país no fator 8	96
Tabela 26 - Discursos presentes no polo positivo do fator 9	97
Tabela 27 - Variáveis discursivas do polo negativo do fator 9	104
Tabela 28 - Escores médios de cada país no fator 9	105

Sumário

INTRODUÇÃO	14
A LINGUÍSTICA APLICADA E A CIÊNCIA ABERTA	19
JUSTIFICATIVA, QUESTÕES DE PESQUISA E OBJETIVOS	21
1. FUNDAMENTAÇÃO TEÓRICA	23
1.1 Linguística de Corpus	23
1.2 Representação e multimodalidade	29
1.3 Análise Multidimensional	30
2. METODOLOGIA	32
3. RESULTADOS	39
3.1 Fator 1: Representações da estética do bélico em coexistência com a do sensível	40
3.2 A beleza natural litorânea frente às paisagens artificiais (Fator 2)	51
3.3 O registro topográfico e o simbolismo político nacional (Fator 3)	59
3.4 O ambiente urbano e os pequenos seres (Fator 4)	66
3.5 A natureza manipulada e fértil (Fator 5)	74
3.6 Relíquias de valor simbólico, icônico ou místico (Fator 6)	79
3.7 A estética em atividades de lazer e dos aparatos de grande escala (Fator 7) 84	
3.8 A tecnologia da engenharia e das edificações (Fator 8)	90
3.9 A vegetação e o olhar turístico-fotográfico (Fator 9)	96
4. DISCUSSÃO	106
5. CONCLUSÃO	111
REFERÊNCIAS	115
ANEXOS	121

INTRODUÇÃO

O interesse pela área de análise imagética começou ainda na minha graduação em Comunicação Social, na PUC-SP, quando realizei pesquisa de iniciação científica, em que me baseei no tripé Comunicação Social, História da Arte e Tecnologia, para entender as questões ligadas à recepção visual. Mais estritamente, o interesse da pesquisa ali realizada foi verificar se os objetivos comunicacionais do autor da mensagem visual, em sua etapa de esboço, eram, de fato, notados pelo observador, no ato de sua recepção. Para tanto, empregou-se a tecnologia de *eye tracking*, tecnologia esta que hoje já está absorvida por empresas de reconhecimento facial e ocular, como Apple, Samsung e muitas *fintechs*¹. Nessa pesquisa de graduação, busquei verificar tal processo através de uma abordagem semiótica, entendida como o estudo dos signos, incluindo a sua recepção (SANTAELLA, 2002).

No presente estudo, estendo o escopo da pesquisa para o campo da Linguística. Busco entender as representações de "brasilidade", expressas em um grande *corpus* de imagens contemporâneas, utilizando a Linguística de Corpus – mais especificamente, a Análise Multidimensional Lexical, desenvolvida no Grupo de Estudos de Linguística de Corpus (GELC). Essa metodologia alia inteligência artificial (visão por computador) e Análise Multidimensional, a AMD (BERBER SARDINHA; VEIRANO PINTO, 2014, 2019). Tal metodologia permite decompor e anotar automaticamente arquivos de imagem e descobrir os padrões de discursos imagéticos, em um corpus de imagens; possibilita também a identificação de sentidos imanentes desses padrões. Por meio dessa análise, pretendo identificar os discursos indexados (ou expressos, presentes) nas imagens, focalizando naqueles que embasam e são tradicionalmente vinculados à ideia dessas representações de Brasil.

Cabe dizer, antes do que se possa vir a acreditar, por razões do nosso tempo presente, que esta não é uma pesquisa contendo ideias nacionalistas, mas

¹ *fintechs* é uma abreviação para *financial technology* (tecnologia financeira) e é usada para se referir a startups ou empresas que desenvolvem produtos financeiros totalmente digitais, nas quais o uso da tecnologia é o principal diferencial em relação às empresas tradicionais do setor.

sim de problematização das repercussões imagéticas da cultura de um povo. As representações de uma cultura estão fortemente afirmadas pelas imagens, construindo impressões e expressões coletivas sobre a cultura. Observa-se que tão logo que as terras em redor do globo foram conquistadas/invadidas pelas metrópoles, a arte visual passou também por uma transformação. Desenvolveu-se o uso da comunicação descritiva e imagética dos "paraísos" encontrados. Surgia a missão dos pintores viajantes (GOMBRICH, 2013). Nesse novo estilo, o autor-emissor das imagens criava mais do que retratava, nas representações das terras e culturas sobre as quais pintava, direcionando assim a percepção dos observadores. Muito se fala de apagamento cultural e estranhamento social nos conteúdos dessa época. A premissa é de que os discursos dominantes criam representações de uma cultura e de uma sociedade. As representações de "Brasil" e "brasilidade" começaram a ser criadas ali e isso ganhou continuidade tradicional. Essa prática perdura até os dias de hoje, na publicidade, por exemplo. Podemos então nos perguntar se até que ponto os agentes originais desse discurso (metrópoles, os artistas europeus, seus mecenas) atribuíram características "não evoluídas", ou por vezes distorcidas e animais, aos habitantes dessas terras distantes. Esse movimento dos pintores viajantes tem recebido interesse do público e de curadores de arte, como na seguinte apresentação da Enciclopédia Itaú:

De modo geral, esses artistas integram expedições artísticas e científicas que, nas Américas, desde sua descoberta, no século XVI, atravessam os territórios recém-conquistados, com a finalidade de registrar a flora, a fauna e seu povos. No caso do Brasil, vastas literatura e iconografia são produzidas desde a chegada dos portugueses no século XVI até o século XIX: os relatos e registros pictóricos descrevem as novas paisagens projetando imagens variadas da terra e do homem. Espécimes naturais desconhecidos, animais estranhos e homens "primitivos" (às vezes "bons selvagens", outras, "selvagens-canibais") compõem o imaginário europeu acerca do Novo Mundo, descrito ora como "inferno", ora como "paraíso terreal". A riqueza da produção dos artistas viajantes - seja pelo seu valor artístico, seja por conta de seus pontos

de vista e suas descrições acerca das novas terras e gentes - desperta a atenção de analistas de diversas áreas: geógrafos, antropólogos, historiadores da arte e da cultura. Os primeiros relatos sobre o Brasil, como o do viajante Hans Staden (1510-1576), vêm acompanhados de ilustrações (em geral, de autoria desconhecida) que traduzem as peripécias do viajante alemão, prisioneiro dos índios tupinambás, ameaçado de morte e canibalismo... E os do pastor calvinista Jean de Léry (1536-1613). O governo de Maurício de Nassau em Pernambuco, de 1636 a 1645, é responsável por fontes iconográficas fundamentais a respeito do Brasil holandês. Entre os artistas que documentam o país nesse contexto encontram-se os holandeses Albert Eckhout (ca.1610-ca.1666) e Frans Post (1612-1680), contratados para integrar a comitiva de Nassau. (ENCICLOPÉDIA ITAÚ, 2020).²

Nesse âmbito, um livro exemplar da missão de Maurício de Nassau é a obra *História Natural do Brasil (Historia Naturalis Brasiliæ)*, de Willem Piso. As imagens ali dispostas, muitas por meio do discurso não-verbal das gravuras e pinturas, passaram a construir o significado de “brasilidade”. O mesmo ocorreu com os recursos verbais. *Historia Natvralis Brasiliae* contém diversas ilustrações, mas é um livro realmente textual. Também como outro exemplo, por sua vez, a carta de Vaz de Caminha já contribuía para a narrativa de paraíso na terra:

E então se começaram de chegar muitos; e traziam cabaças d'água, e tomavam alguns barris que nós levávamos e enchiam-nos de água e traziam-nos aos batéis. E nós tomávamo-los. Avistamos três ou quatro Papagaios vermelhos, muito grandes e formosos, e dois verdes pequeninos... Nos traziam carapuças de penas verdes, e um pano de penas de muitas cores, espécie de tecido assaz belo...esta gente é boa e de bela simplicidade. E imprimir-se-á facilmente neles qualquer cunho que lhe quiserem dar, uma

² Verbete *Pintores Viajantes*, trecho extraído da Enciclopédia Itaú Cultural. ARTISTAS Viajantes. In: ENCICLOPÉDIA Itaú Cultural de Arte e Cultura Brasileira. São Paulo: Itaú Cultural, 2023.

vez que Nosso Senhor lhes deu bons corpos e bons rostos, como a homens bons. E Ele nos para aqui trazer creio que não foi sem causa. E portanto Vossa Alteza, deve cuidar da salvação deles. E prazera a Deus que com pouco trabalho seja assim! Eles não lavram nem criam. Nem há aqui boi ou vaca, cabra, ovelha ou galinha, ou qualquer outro animal que esteja acostumado ao viver do homem. E não comem senão deste inhame, de que aqui há muito, e dessas sementes e frutos que a terra e as árvores de si deitam. Nesse dia, enquanto ali andavam, dançaram e bailaram sempre com os nossos, ao som de um tamboril nosso, como se fossem mais amigos nossos do que nós seus... Até agora não pudemos saber se há ouro ou prata nesta terra de fartura, ou outra coisa de metal, ou ferro; nem lha vimos. Águas são muitas; infinitas. Em tal maneira é graciosa que, querendo-a aproveitar, dar-se-á nela tudo; por causa das águas que tem! Contudo, o melhor fruto que dela se pode tirar parece-me que será salvar esta gente (CAMINHA, 1500).³

A citação deixa evidente um discurso de apreciação e função estética a serviço da finalidade colonial. Nascia o vestígio mitológico de brasilidade. Em outros trechos, aos indígenas nem sequer foi atribuída a ideia de povo, mas sim de gente com potencial de se tornar um povo, "como homens bons", com a "simplicidade de Adão" e "inocência". Fato marcante é a construção estética de paraíso na terra e que tal estética ainda é anunciada e mantida atualmente, principalmente por meio de imagens veiculadas na web e na cultura visual brasileira.

³ Carta de Pêro Vaz de Caminha. 1 de maio de 1500. Portugal, Torre do Tombo, Gavetas, Gav. 15, mç. 8, n.º 2.

Assim, "brasilidade" é a palavra para a qual eu devo o tema e a ideia deste trabalho de pesquisa, baseando-me em Berber Sardinha (2017), que, por sua vez, emprestou a ideia de Barthes (1977). Aplicada quase sempre como um rótulo, o termo serve de adjetivo ou sinônimo àqueles atributos os quais traduzem anos de alteridade e da impressão indelével que identifica a imagem brasileira. Esta pesquisa é constituída, sobretudo, a partir da Linguística de Corpus, na metodologia da análise multidimensional (AMD), o que permite uma análise a respeito da construção do que representa essa 'brasilidade' num grande volume de dados.

Inicialmente, ao adentrar no estudo dos principais paradigmas que precederam esta pesquisa, e objetivando situá-la perante os mais variados conceitos históricos, pode-se dizer que a metodologia aqui utilizada é fruto de uma *bricolage* (MOITA LOPES, 2006, p.101), que vem a ser 'a união da teoria e prática em uma formulação do conhecimento, levando em consideração os contextos históricos e sociais daqueles que vivem o conhecimento na prática'.

Além disso, como investigação científica no campo da Linguística Aplicada, a pesquisa busca não só evocar, nas linhas deste trabalho, reflexões que porventura possam contribuir minimamente para o avanço acadêmico da área, mas que também possibilitem ou instiguem alguma transformação em nossa sociedade (PENNYCOOK, 2006), tal como entender o que vem a ser a 'brasilidade' e expor metodologias científicas em ciência aberta. Conforme colocado, esta pesquisa está inserida na área da Linguística de Corpus, visto que emprega elementos teóricos e metodológicos da área, tal como trabalhados no Grupo de Estudos em Linguística de Corpus - GELC, por meio da extensão da Linguística de Corpus como área que busca a descrição de coletâneas de imagens, além de textos escritos e falados, como tem se caracterizado desde os anos 1960. Desse modo, para o GELC, ao assumir a tarefa de dar conta da descrição de corpora de imagens, a Linguística de Corpus faz uma intersecção com as Humanidades Digitais (MANOVICH, 2011; NYHAN; PASSAROTTI, 2019), no sentido em que empresta o objeto de estudo (bibliotecas de imagens) e as interpela com métodos oriundos da Linguística (BERBER SARDINHA, 2020).

Posterior a uma revisão panorâmica das variações conceituais e epistemológicas da Linguística Aplicada, cabe o apontamento de que algumas posturas (ou afiliações políticas) estão presentes no decorrer da análise aqui

proposta. Primeiro, toma-se neste trabalho uma visão decolonial, decorrente da teoria crítica da linguagem por Costa (1980), Carneiro (2006) e Fanon (2002). Segundo, há nela a adoção de uma visão pós-positivista e humanista, possibilitando entender a ciência e o fazer pesquisa em humanidades não como um ato preciso e engessado em dogmas ou verdades, mas como uma das inúmeras soluções possíveis frente à outras várias questões em potencialidades (FANON, 2002). Isso devido à consciência do quanto o ser humano enquanto ser de peculiaridades e com isso, por ser uma criatura mutável, suas percepções, conclusões, noções científicas e posturas também o podem ser. Posturas essas a serem lidas criticamente, uma vez que mesmo as epistemologias são frutos de estruturas de dominação e exercício de poder sobre os sujeitos (MORIN, 2005). Por fim, em terceiro lugar, a pesquisa que aqui se desenvolve apresenta-se como inter e transdisciplinar. Interdisciplinar porque a investigação executou-se na linguística, com a linha e ferramentas metodológicas da Linguística de Corpus (BERBER SARDINHA, 2004). Porém, combina também as disciplinas da visualidade e da sociologia e ciências políticas. E transdisciplinar por agregar algumas "bricolagens" metodológicas. Por exemplo, a interface do Google Vision Cloud sendo usada para etiquetagem em Linguística de Corpus e a pesquisa em linguística debruçada não sobre textos e palavras, mas exclusivamente sobre fotografias.

A Linguística Aplicada e a Ciência Aberta

Fez parte do desenvolvimento desta pesquisa a construção do Portal Multimodal de Ciência Aberta para o Avanço das Humanidades, ou simplesmente o Portal de Ciência Aberta (www.cienciaaberta.org), projeto de onde o autor recebeu o financiamento do CNPq. Essa atividade do pesquisador esteve relacionada diretamente a design gráfico, design web, edição de imagem e vídeo, publicação digital e aplicação de UX design⁴. Isso em termos meramente operacionais. Pois em termos de produção do saber, trata-se de adesão a um novo modo e paradigma de se fazer uma pesquisa. Durante todo o andamento do trabalho de pesquisa

⁴ Design de experiência do usuário.

realizou-se a publicação dos avanços, percalços e ajustes, no formato de postagens-diário, ainda que espaçadamente, no blog do portal. Ou seja, o caderno de anotações, com toda a jornada do pesquisador, esteve à disposição do público em geral e continua no ar, no Portal, ao livre acesso à web⁵.

Desenvolver pesquisa nos moldes e pilares da ciência aberta nas humanidades trouxe para o desenvolvimento dos estudos das Humanidades, e no presente caso, para a pesquisa em linguística aplicada, a possibilidade de se realizar um discurso apologético, necessário na medida em que a cientificidade dos estudos das ciências humanas esteve posto em xeque nos últimos anos. Questionava-se sua real entrega tangível, quantitativa, mercadológica enquanto *produto* para a sociedade. Obviamente, ideias de instrumentalização do saber, oriundas de um neoliberalismo global e incoerentes com o propósito da academia, que é a pesquisa, o ensino e a extensão universitária. Assim, tal apologética se fez possível, primeiro, graças ao esclarecimento de quais tipos de resultados as humanidades se ocupam, a saber, os conceitos epistemológicos. Conceitos ligados ao comportamento humano e seu histórico ou contextos. Como nas ciências naturais, onde os conceitos e pedras fundamentais são semânticos e matemáticos, físicos, químicos ou biológicos, todos com suas leis, nas humanidades a maior lei seria a responsabilidade de interpretação do *zeitgeist*⁶; Segundo, tal apologética só faz sentido mediante uma postura consciente da responsabilidade do impacto social perante os dados manuseados durante um estudo e mediante a disponibilização destes dados à comunidade, como verdadeiros bens e frutos contínuos, do trabalho de pesquisa em ciências humanas. No entanto, parte desse ataque e ideia de inutilidade social das linhas de pesquisa em humanidades é também advinda da própria postura tradicional, positivista, universalista e de fechamento acadêmico ao mundo exterior, não permitindo a troca ou a disponibilização de discussões com os *stakeholders* sociais e políticos. Por isso, nesse novo paradigma o dado não é uma propriedade. Pelo contrário, "o dado é dado" (BERBER SARDINHA, 2022); E em terceiro lugar, a validade da pesquisa e a consistência dos achados em ciências humanas ganham a solidez com a prática

⁵ Ver exemplos das postagens nos Anexos.

⁶ Ou "espírito do tempo".

constante da revisão por pares, algo possível apenas com a livre circulação das pesquisas. O escrutínio é uma premissa na ciência aberta. O resultado deste paradigma é a maior assertividade e revisão constante, em busca das conclusões mais concretas e da revisão crítica do estado da arte, preenchendo lacunas.

Face a essa problemática, o presente trabalho teve por objetivo geral contribuir para a criação e alimentação do Portal Multimodal/Multilíngue para o Avanço da Ciência Aberta nas Humanidades e, como objetivo específico, constituir um segmento, de sua base de dados, voltado para a produção e análise de práticas intersemióticas e discursivas que circulam na contemporaneidade, como as questões ligadas à brasilidade.

Essa nova aurora científica transformou muito o fazer científico já neste trabalho. E agora, sem mais delongas, o estudo em si, com suas análises, discussões e conclusões, se apresenta a seguir, a começar pelas questões da investigação.

Justificativa, questões de pesquisa e objetivos

Apesar da abundância de imagens que indexam o Brasil e a brasilidade em circulação na web (e também nas redes sociais, livros impressos, portfólios online, entre outros), e já posteriormente a um rico estado da arte ligado aos estudos da mensagem semiótico-visual, da filosofia da imagem e da imagem do Brasil em trabalhos de diplomacia e turismo⁷, não existem estudos de larga escala de dados, computacionais-imagéticos, sobre quais representações surgem do Brasil, estritamente; Nem se há variação em relação a representações de outras culturas nacionais que emergem das imagens em circulação. Até onde se pode verificar, este seria o primeiro estudo de *corpus* visual sobre representações imagéticas de brasilidade.

⁷ Bakhtin (1963:2010) e o círculo, aponta Brait (2005), foi quem mais extrapolou as análises dos sentidos dos enunciados para além do textual, com sua perspectiva metalinguística e dialógica; Barthes (1977), Kress e Van Leeuwen (2001) e Bateman (2011) foram grandes investigadores da filosofia da imagem, no campo linguístico; Da mesma maneira, é importante herança as pesquisas nos campos da imagem, enquanto propaganda discursiva, que Kajihara (2008) e Gadelha (2018; 2020) realizaram, com estudos de compilação e análise manual da imagem internacional do Brasil.

Do mesmo modo, há poucos estudos sobre representações do Brasil em textos escritos. Berber Sardinha (2019) realizou uma pesquisa sobre representações do Brasil em publicações do Google Books, cobrindo um período de mais de 200 anos de produção textual. Esse foi um estudo de linguística de corpus na perspectiva de 'big data', que se baseou em uma massa de dados de quase um bilhão de palavras. As representações de brasilidade identificadas pelo autor no referido trabalho foram descritas em torno de cinco 'dimensões', sendo a primeira 'a economia e a política', a segunda 'a arte tradicional, ciências, as pessoas e a terra', a terceira 'matéria prima e a paisagem', a quarta 'novas formas artísticas, mulheres e homens, religião e o meio ambiente' e a quinta 'a monarquia, transporte a vapor e a região selvagem'.

Assim, a lacuna que este estudo se prontifica preencher na literatura refere-se à análise das representações imagéticas de 'brasilidade', na perspectiva da Linguística de Corpus, isto é, que utilize grandes quantidades de dados pictóricos. A fim de ajudar a preencher essa lacuna, o objetivo geral do presente projeto é descrever as representações de brasilidade que emergem de um grande *corpus* de imagens referentes ao Brasil. Dessa forma, este projeto pretende contribuir tanto para a expansão dos estudos sobre brasilidade com base em grandes *corpora*, quanto para impulsionar os estudos em historiografia da arte, focados na análise de dados visuais, isto é, imagens históricas (pinturas, gravuras, desenhos, ilustrações) e contemporâneas (figuras, memes, dentre outros) veiculadas na web.

Para o presente estudo foi compilado um *corpus*, com grande número de textos visuais, relacionados à representação temática de Brasil e brasilidade. O desenho do *corpus* está detalhado na seção de metodologia.

As inovações ou potencialidades que pautaram essa pesquisa incluem:

- Contribuir para o avanço da Linguística de Corpus, ajudando a operacionalizar pesquisas sobre análise de *corpora* visuais;
- Contribuir para a interface da Linguística de Corpus com a Análise do Discurso, identificando a presença de representações de brasilidade em *corpora* visuais;

- Empregar novas tecnologias de análise automática de imagens, com recursos computacionais e de Inteligência Artificial, como o Google Cloud AI;
- Compreender a variação das representações de Brasil e brasilidade diacronicamente, desde as primeiras pinturas e gravuras sobre o Brasil até as imagens que circulam na web nos dias atuais.

A pesquisa pretende contribuir para um melhor entendimento de como o Brasil tem sido representado historicamente e na contemporaneidade por meio de imagens. A pesquisa mostrará quais as principais características das imagens e qual a variação entre as imagens ao longo do tempo, por comparação discursiva e geográfica (Brasil, Estados Unidos, Reino Unido e Alemanha).

A questão de pesquisa que permeou o desenvolvimento deste trabalho é empírica, como próprio da investigação em linguística de corpus, qual seja:

- Quais os discursos e como o Brasil é representado num *corpus* visual e em que medida esses possíveis discursos divergem do senso comum do que significa?

1. FUNDAMENTAÇÃO TEÓRICA

1.1 Linguística de Corpus

Esta pesquisa está inserida na Linguística Aplicada. Mais especificamente, em Linguística de Corpus, ou como é abreviada, a LC. A LC é uma área de pesquisa da linguística que atua e lança mão à computação, para o fazer científico. Nesse sentido, seus primeiros feitos e estudos datam de 1946, com estudos de lematização e indexação computadorizada, sendo os primeiros deles *The Index Thomisticus* e o *The Brown Corpus*, no momento em que a computação veio somar às pesquisas de humanidades, mais precisamente, à linguística (NYHAN;

PASSAROTTI, 2019). Tal encontro passou, aliás, a ser chamado de "humanidades digitais" (NYHAN; PASSAROTTI, 2019; BERBER SARDINHA, 2022).

O principal referencial teórico da pesquisa aqui apresentada é a herança epistemológica da Linguística de Corpus, que é entendida como uma área dos Estudos Linguísticos, que se utiliza de grandes quantidades de textos, sejam eles falados, escritos, visuais ou acústicos, a fim de detectar padrões de uso (BERBER SARDINHA, 2004). Inicialmente, a Linguística de Corpus foi introduzida para trabalhar com *corpora* monomodais, textuais, nos anos 1960. As pesquisas realizadas desde então enfocaram o uso linguístico, tais como padrões de colocação, variedade lexical, aquisição de língua, coligação e prosódia semântica (BERBER SARDINHA, 2004).

Faz parte dos seus procedimentos, antes de mais nada, a noção de *corpus*, ou *corpora*, plural do primeiro. Trata-se do uso e compilação de "um grande volume de dados" a fim de responder a uma questão linguística (BERBER SARDINHA, 2004). Noutras palavras, para se fazer uma pesquisa em LC sempre haverá um banco de dados, de textos, que é processado, lido e analisado por computador, a serviço da pesquisa. Atualmente, devido a maior capacidade de computação, somados à difusão da LC e ao advento da inteligência artificial, o volume de textos e de pesquisas nesse campo da linguística tem crescido.

Podemos dizer que desde sua idealização nos anos 1940 até o presente, há no mundo três escolas de linguística de corpus: A escola britânica, com destaque para a Universidade de Lancaster; A escola americana, com trabalhos expressivos na universidade de Flagstaff (Northern Arizona University - NAU), aprimorada em torno do pensamento de Douglas Biber; e a escola brasileira, de São Paulo (PUC-SP), que criativamente desenvolveu suas características e metodologias empíricas e probabilísticas próprias, a partir da união do pensamento das duas escolas supracitadas, sob a supervisão do Dr. Tony Berber Sardinha.

A Linguística de Corpus passou a ser usada mais recentemente na interface com a Análise de Discurso, sendo empregada para detectar os discursos correntes em determinados *domínios* de uso da língua. Domínio aqui entende-se como grande campo temático ou do saber, representado no corpus em análise. Dessa

maneira, cabe destacar que é de suma relevância nesse tipo de pesquisa a observância de que a compilação de um *corpus* possa atingir certos critérios explicitados pelo pesquisador ao coletar os textos ou imagens (como na presente pesquisa). Seriam esses critérios uma delimitação de domínio, identificação da modalidade, especificação da cronologia, seleção de como será composto, e uma descrição de seu conteúdo (ROMEIRO, 2020; BERBER SARDINHA, 2004), além das metodologias estatísticas e de pesquisa por amostragem de universo.⁸

Berber Sardinha (2019) empregou a Linguística de Corpus para detectar os discursos responsáveis pelas representações históricas do Brasil e dos EUA. Nesse tópico de sua obra, utilizando um corpus de mais de 500 milhões de palavras, disponibilizado pelo Google Books, o autor identificou as principais construções linguísticas que conjuntamente construíram as marcas de ‘brasilidade’ e de ‘americanidade’.

As representações de brasilidade identificadas pelo autor no referido trabalho foram descritas em torno de cinco ‘dimensões’:

1. A economia e a política
2. Arte tradicional, ciências, as pessoas e a terra
3. Matéria prima e a paisagem
4. Novas formas artísticas, mulheres e homens, religião e o meio ambiente
5. A monarquia, o transporte a vapor e a região selvagem
(BERBER SARDINHA, 2019, p. 255)

As evidências linguísticas que permitiram a identificação dessas representações foram os bigramas (sequências de duas palavras) presentes nas publicações do Google Books. Foram selecionados os bigramas que começavam com a palavra ‘*Brazilian*’, no caso das representações do Brasil, e ‘*American*’, no caso das representações dos EUA. Para ilustrar, os termos que se colocavam com ‘*Brazilian*’ nos bigramas, referentes à primeira representação apontada acima (‘a

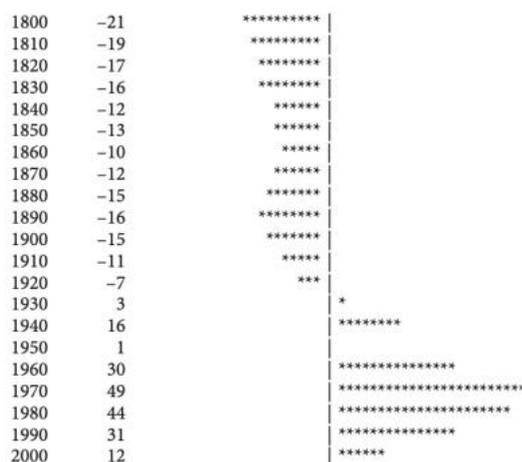
⁸ Para melhor compreensão e explicitação da aplicação desses critérios na presente pesquisa, ver tabela no item 1, “Design de Corpus”, na seção Metodologia.

economia e a política), foram termos como ‘economia’, ‘desenvolvimento’ ‘industrial’, ‘experiência’, ‘militar’, ‘político’, ‘firmas’, ‘caso’, ‘exportar’, entre outras. (BERBER SARDINHA, 2019, p. 248)

A partir da análise desses termos, dos discursos que eles indexam, e da análise qualitativa de obras e período em que os bigramas mais ocorreram, o autor pôde propor que esse conjunto lexical apontava para a representação do Brasil como ‘economia’ e como ‘política’, isto é, um país do qual há um interesse em salientar aspectos econômicos e políticos.

Para a identificação das representações, foi muito importante a verificação dos períodos históricos em que ocorreram, pois os discursos são situados sócio-historicamente. Especialmente levando-se em conta que os textos do Google Books foram escritos em língua inglesa, geralmente por autores norte-americanos e britânicos, do sexo masculino e brancos. O gráfico abaixo mostra a distribuição do primeiro discurso ao longo do tempo:

Figura 1 - Incidência média da dimensão 1 no Google Books Corpus.



Fonte: BERBER SARDINHA, 2019, p. 249

O gráfico mostra que a representação do Brasil como tendo uma economia importante e um cenário político saliente acontece principalmente a partir dos anos 1960. Isso coincide com o período do milagre econômico brasileiro, quando a

economia do país crescia dois dígitos em média por ano, despertando o interesse do público internacional. O gráfico mostra ainda que essa representação praticamente não existia por mais de 100 anos, ou seja, a imagem do Brasil não foi, por um longo período, a de uma sociedade que tinha economia ou cenário político notável.

Apesar de não descrever as demais dimensões identificadas, é importante salientar que o estudo mostra a possibilidade de detectar representações de uma determinada cultura por meio da análise de grandes quantidades de dados linguísticos. Normalmente, os estudos sobre representações sociais ou culturais, em pesquisas de universo limitado e entrevistas com *focus group*, por exemplo (GATTI, 2005; GREENBAUM, 1998) empregam pouca quantidade de dados, que são analisados qualitativamente, mas com critérios de representatividade. Não se trata de defender uma ou uma outra abordagem como sendo ‘superior’, mas sim de mostrar que é possível inovar ao usar a tecnologia para identificar construtos discursivos em larga escala.

Na presente pesquisa, utilizamos as seguintes definições de ‘discurso’, que foram empregadas por Berber Sardinha (2019) em seus estudos na interface da Linguística de Corpus com Análise de Discurso. Segundo Hajer (1995, p. 44), discurso é:

Um conjunto de ideias, conceitos, e categorizações que são produzidas, reproduzidas e transformadas em um determinado conjunto de práticas pelas quais o significado é dado às realidades físicas e sociais (HAJER, 1995).

Ou seja, seguindo essa concepção de discurso, podemos dizer que os discursos são construções abstratas que se transformam em práticas sociais, que por sua vez guiam o entendimento dos indivíduos. Os discursos possuem, na verdade, um poder de enquadramento da realidade, fazendo um recorte de uma porção do contexto, ao mesmo tempo em que esconde outra porção: “Os discursos dão forma a certos problemas; ou seja, eles distinguem alguns aspectos de uma situação em detrimento de outras” (HAJER, 1993, p. 45).

Outra definição de discurso em que nos apoiaremos é a de Hall (1992), segundo a qual os discursos se apresentam na superfície muitas vezes como modos de tratar determinados ‘assuntos’:

um grupo de afirmações que fornecem uma língua sobre a qual se fala - por exemplo uma maneira de representação - de um determinado tipo de conhecimento sobre um tópico. Quando afirmações sobre um tópico são feitas dentro de um discurso em particular, o discurso torna possível a construção do tópico de uma certa maneira. Ela também limita outras maneiras de que o tópico pode ser construído (HALL, 1992, p. 92).

Desse modo, segundo Berber Sardinha (2019), na perspectiva da Linguística de Corpus, os indícios de discursos podem ser ‘mineirados’, isto é, extraídos do corpus pelo computador, por meio do léxico, o qual aponta os possíveis ‘assuntos’, ou melhor, discursos que estão sendo tratados nos textos. De posse dos indícios identificados pelo computador, o analista pode interpretá-los como discursos e como representações, após fazer análise qualitativa de textos onde esses indícios ocorrem.

Ao mesmo tempo em que Berber Sardinha (2019) mostra a possibilidade de identificar representações nacionais em larga escala, seu estudo restringiu-se à modalidade verbal escrita. No entanto, representações são criadas e reproduzidas muito frequentemente por meio da modalidade visual.

Assim, no presente projeto, utilizou-se os construtos empregados por Berber Sardinha (2019, 2020) para a análise de representações em um corpus monomodal, imagético, para identificar as representações visuais de brasilidade. Para permitir essa identificação, será preciso coletar um grande corpus de imagens e realizar a análise automática dessas imagens por meio de ferramentas de Inteligência Artificial, que neste estudo será o Google Cloud Vision AI. A coleta do corpus e a operação dessa ferramenta será detalhada na metodologia.

1.2 Representação e multimodalidade

Campos como a linguística e a filosofia, dentre outros, ocupam-se da ideia de representação. Muito se aborda sobre a noção de vontade em Schopenhauer, por exemplo, mas deve-se atentar também para como esse autor pensava a representação. Em *O mundo como vontade e representação*, esta última é defendida como o real atributo de tudo aquilo que nos cerca. Noutras palavras, é interessante para compreensão de uma análise imagética a reflexão sobre o que seria o mundo e tudo que há nele, enquanto uma representação de algo, uma ideia e porque não dizer, um discurso.

Esta mesma interpretação nos chega, nas escolas linguísticas, por intermédio de duas abordagens da filosofia aplicada ao estudo da cultura e das linguagens: a abordagem do discurso e a abordagem da semiótica. No trabalho intitulado "*Representação e Cultura*", Hall (2016) identifica que nas representações de um povo, há um sistema que pode ser chamado de "circuito da cultura". É o mesmo que um imaginário ou uma ideologia de reconhecimento e de identificação social. É uma linguagem que opera no que o autor denominou "sistema representacional" (HALL, 2016). Nesse contexto, cultura "é aquilo que melhor foi dito ou pensado por um povo", ideias questionadas por Hall, principalmente por serem justificativas da distinção de uma "cultura elevada" e uma "cultura popular". No entanto, nessa concepção, dizer que dois indivíduos têm a mesma cultura é dizer que eles veem o mundo da mesma maneira. Para o autor, a cultura é um código complexo, interpessoal, semiótico e visual, entre os povos. Entre os sujeitos, a representação é um "ativo" a ser zelado, porém é volátil, de acordo com os valores do contexto temporal, isto é, da cultura (HALL, 2016).

Como dito anteriormente, Hall discutiu a questão de discursos visuais (HALL, 2016), entendendo o discurso, ou a propaganda de uma cultura, como algo muito mais imagético do que verdadeiramente dito. Para isso, o autor se baseia nos pensamentos de Foucault, Barthes, Peirce e Saussure.

Há também os esforços da análise iconográfica, desenvolvida por Aby Warburg e Erwin Panofsky (2018), a respeito de como ciência da história da arte se preocupa em extrair discurso, significado e sentidos das imagens artísticas. Tal metodologia é reconhecida por exigir uma imersão no discurso a ser analisado, exigindo um grande repertório na interpretação dos enunciados e seus contextos, isoladamente, caso a caso.

No entanto, em ambos os casos, a leitura das imagens e seus significados têm limitações quantitativas. É possível analisar uma imagem por vez. Nesse sentido, a pesquisa empírica de linguística de corpus está aqui proposta como um avanço relevante também para os estudos das imagens. Para tanto, lançou mão do estudo da multimodalidade, associado à ideia de discurso monomodal de Bednarek (2018), e os aplicou às imagens, em uma LC multimodal (DELFINO, 2022; BATEMAN, 2011; BERBER SARDINHA, 2019).

1.3 Análise Multidimensional

Entre os instrumentos de análise à disposição da linguística de corpus, a análise multivariada denominada Análise Multidimensional (BIBER, 1988; BERBER SARDINHA; VEIRANO PINTO, 2014, 2019) foi a selecionada para esta pesquisa, na medida em que esta técnica investiga elementos subjacentes à superfície linguística formada pelas palavras organizadas no sistema da léxico-gramática, para descrever discursos, seus registros, gêneros e estilos (BIBER; CONRAD, 2009).

A Análise Multidimensional (AMD) possibilita observar um texto, ou grupo de textos homogêneos, conforme sua classificação atribuída socialmente, através de diferentes dimensões. Por sua vez, tais dimensões sintetizam atribuições linguísticas comuns (variáveis), presentes ou ausentes no texto, que são interpretadas na forma de um contínuo que não admite opções binárias, apesar de existir por vezes polarização entre os componentes de cada dimensão, dado que as variáveis da dimensão podem ser positivas ou negativas (BROGINI, 2023). Na Análise Multidimensional originalmente desenvolvida (BIBER, 1988), a língua é analisada em sua função. Nela, o objetivo é criar um modelo no qual existem

grupos de variáveis concomitantes, cuja variação entre textos pode ser alvo de uma interpretação detalhada e de ordem qualitativa.

A ideia de extração de grupos de variáveis para expressar dimensões latentes de variação se estende a outros tipos de Análise Multidimensional desenvolvidos após a publicação de Biber (1988), que utilizam não mais categorias gramaticais, como também de categorias do léxico e da semântica (BERBER SARDINHA; VEIRANO PINTO, 2012; 2019; BERBER SARDINHA, 2020). Segundo Berber Sardinha (2021), a meta de uma Análise Multidimensional Lexical é revelar discursos subjacentes à superfície linguística de muitos textos, vistos sob uma análise multivariada:

[O] objetivo principal de uma análise de discurso multidimensional é descrever a variação do discurso; ou seja, como o discurso varia sistematicamente de acordo com o contexto. A variação é modelada a partir de uma perspectiva multidimensional, significando que cada texto é visto como sendo formado simultaneamente pela incidência dos vários discursos representados por várias dimensões.

(BERBER SARDINHA, 2021, p. 298-299)

Desse modo, definimos a utilização da Análise Multidimensional Lexical neste estudo a partir de suas variáveis de natureza lexical: há os lemas agrupados por categorias gramaticais, como substantivos, adjetivos e verbos. Isso significa a junção de formas flexionadas nominais e verbais, que estão reunidas na forma canônica dos lemas.

A seguir, apontamos o passo a passo para a realização da AMD lexical nesse estudo, segundo Berber Sardinha (2014; 2017):

1. Identificação e contagem das palavras;
2. Normalização das frequências das variáveis lexicais;
3. Extração fatorial inicial não rotacionada baseada nas frequências normalizadas para identificar os fatores a serem utilizados;

4. Scree plot: Definição do número de fatores para análise por meio de um gráfico de análise de sedimentação;
5. Eliminação das variáveis lexicais com comunalidades menores que 0,15 (cf. BIBER, 2006, p. 183);
6. Extração fatorial final rotacionada contendo o número de fatores estabelecidos para a análise;
7. Cálculo da quantidade de variação compartilhada pelos fatores extraídos;
8. Checagem da variância dos fatores;
9. Cálculo dos escores de fator de cada texto;
10. Interpretação dos fatores em termos de seus discursos subjacentes por meio da observação dos textos, registros e variáveis.

Diversos estudos de linguística de corpus realizaram suas análises orientados pela Análise Multidimensional Lexical. Comumente, destacam-se Mayer (2018) Romeiro (2020); Kauffmann (2020); Veiga (2021) e Souza (2020). Algumas dessas pesquisas, pautando corpus focados em redes sociais (SHEPHERD; SALIÉS, 2013; BERBER SARDINHA, 2014; 2018). Clarke e Grieve (2019) descreveram a linguagem expressa pelo ex-presidente Donald Trump por meio de um corpus de postagens na rede social Twitter, coletadas entre 2009 e 2018. No estudo, que usou um modelo de análise fatorial sobre etiquetas de categorias gramaticais, foram identificados "estilos de comunicação" com discursos trumpistas mais (ou menos) conversacionais (CLARKE; GRIEVE, 2019).

2. METODOLOGIA

Conforme mencionado, este estudo emprega um corpus de imagens relativas ao Brasil, compilado para ser uma amostra criteriosa e representativa das imagens em circulação sobre o Brasil:

Por "criteriosa" entende-se que a amostra deve refletir a variedade escolhida o mais fielmente possível. Além de ser

compatível com os objetivos da pesquisa (Hasan, 1992), a escolha deve ser feita com cuidado, visando a incorporar somente aquele material necessário para representar a amostra que se deseja. Por exemplo, se se quiser construir um corpus geral de uma língua, deve-se fazer uma escolha a mais variada possível: ela deve incluir o maior número possível de registros encontrados na língua-alvo e cada registro, por sua vez, deve ter o maior número possível de exemplares (BERBER SARDINHA, 2000).

Determinante para a pesquisa em LC, o *design de corpus* é uma das etapas mais cruciais para este tipo de pesquisa. Primeiro, porque não é uma escolha arbitrária e ou subjetiva do pesquisador. O motivo disto é também o segundo aspecto aqui destacado, a saber, o *design de corpus* não é apenas uma etapa de pesquisa, mas também um processo, metodologicamente previsto em LC e marcado por critérios próprios.

Um desses critérios é seu atributo de representatividade. Um *corpus* deve representar, com assertividade, o domínio do trabalho em questão, da mesma forma em que uma amostra deve ser fidedigna e precisa para representar todo o universo. Para o presente estudo, a descrição tipológica é a seguinte: modalidade: visual, tempo: contemporâneo, seleção: amostra e conteúdo: especializado.

Tabela 1 - Tipologia do Corpus analisado nesta pesquisa

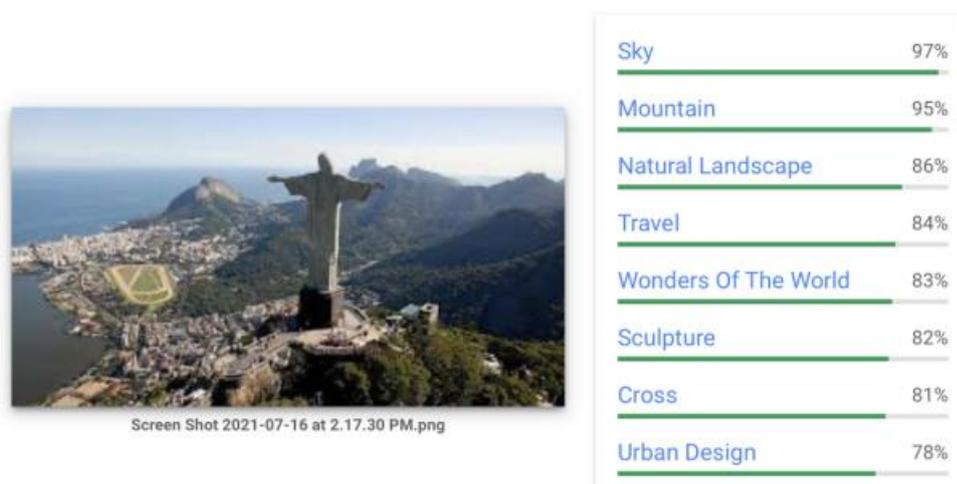
Domínio	Arte
Modalidade	Visual
Tempo	Contemporâneo
Seleção	Amostra
Conteúdo	Especializado

Embora tivéssemos pensado em coletar imagens produzidas dos pintores viajantes, como Johann Rugendas, Eckhout e Debret, até o momento não foi possível, devido a questões práticas. E apesar de muitas dessas peças terem sido enviadas à Europa, algumas seguem no Brasil em acervos digitais, públicos e físicos, como o Instituto Moreira Salles e o Museu Paulista da Universidade de São Paulo. Tais obras foram responsáveis por retratar e “fotografar” registros descritivos de como eram as terras, paisagens e povos daqui.

Porém, conforme mencionado, devido a condições de execução prática, ou seja, não conseguimos coletar nem processar as imagens supracitadas em formato eletrônico, o presente estudo se ateve em analisar imagens produzidas no contemporâneo. Assim, nessa etapa, foi levantado e compilado grande volume de *corpus* (5833 itens fotográficos) do site-portfolio Flickr que, como qualquer imagem presente num site como esse, são itens que possuem um sistema de indexação (tags) pelos próprios autores. Com isso, foi estratégico o aproveitamento dessas tags nativas, separando as imagens com "tags" relacionadas aos países. As tags focalizadas, devido a grande quantidade disponível ali, são: "Brazil", "USA", "UK" e "Germany". Devido também ao fato de sites como esse direcionarem as buscas de acordo com a localização do usuário (endereço de IP) e que esse direcionamento afeta os resultados das buscas, as imagens foram obtidas com o uso da tecnologia VPN, gerando um IP de cada um dos quatro países a seguir: Brasil, Estados Unidos, Reino Unido e Alemanha. O login efetuado por meio de um serviço de VPN, faz com que o computador do usuário esteja de fato conectado a outro, no país desejado. Desse modo, o Flickr (assim como qualquer outro serviço da web), considera que o usuário está naquele país logado via VPN, e não no país onde a máquina do usuário de fato se localiza. Por exemplo, ao logar do Brasil via VPN para um IP do Reino Unido, o Google vai considerar que o computador utilizado está de fato no Reino Unido (embora fisicamente não esteja), retornando assim resultados de busca que enviaria a um usuário britânico. O mesmo acontece com os demais destinos do VPN: os resultados das buscas mostrarão como o Google entregaria imagens sobre o referido país para usuários daqueles países. As imagens foram obtidas automaticamente do Flickr por meio de um script desenvolvido pelo professor orientador.

De posse das imagens coletadas, elas foram etiquetadas pelo **Google Cloud Visual IA**, também por meio de um script desenvolvido pelo professor orientador. Essa ferramenta descreve as características visuais de cada imagem, marcando-as com rótulos (etiquetas) próprios do Google. Esses rótulos são posteriormente colocados em arquivos texto do tipo .JSON, em que as informações são delimitadas por caracteres e códigos especiais. Isso nos possibilita extrair a leitura e compreensão de cada imagem em discursos isonômicos, isto é, não subjetivos, de maneira automática e consistente. Assim, o Google Cloud Vision IA entrega, para cada imagem, um conjunto de descritores textuais, relativos ao reconhecimento, por inteligência artificial, de objetos, paisagens, cores e outras características, conforme mostra a Figura 2 a seguir.

Figura 2 - Output do Google Visual Cloud AI para uma imagem do Cristo Redentor.



Fonte: Captura de tela do Google Cloud Vision.

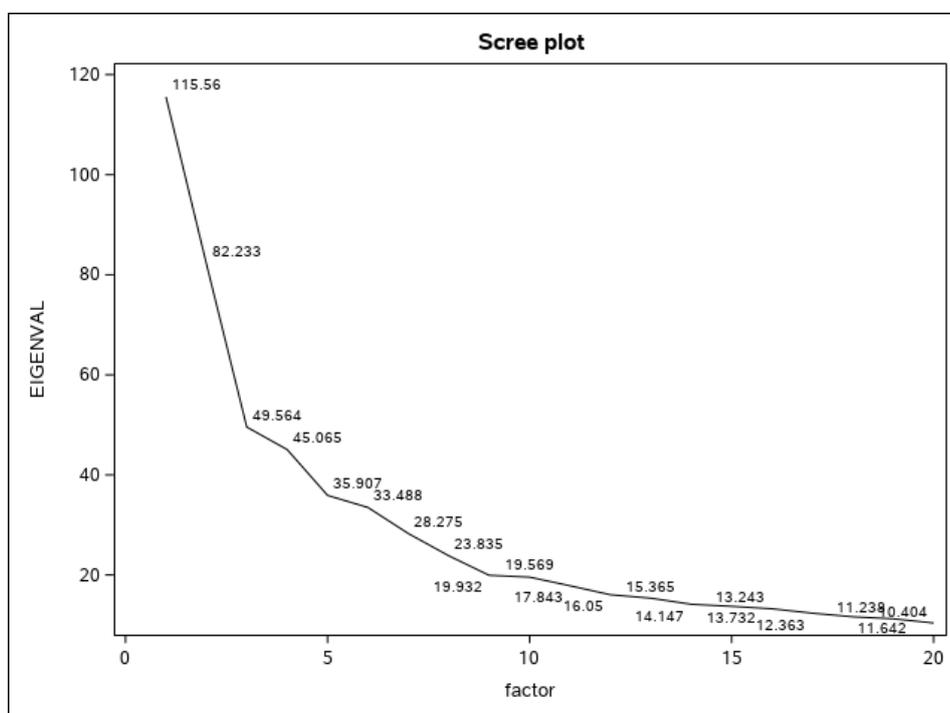
Após as imagens serem etiquetadas automaticamente, as características foram processadas por um script de computador desenvolvido pelo professor orientador, que fez a extração e contagem de todos os rótulos imputados pelo Google Cloud Vision AI.

Em seguida, um outro script desenvolvido pelo professor orientador fez o processamento estatístico das contagens das características visuais, por meio do

pacote estatístico SAS On Demand. A planilha gerada com a contagem dessas variáveis (características visuais) foi o input para gerar a análise fatorial, principal técnica estatística da análise multidimensional.

Para desenvolver uma análise multidimensional, faz-se necessário 02 análises fatoriais, uma não rotacionada, com o objetivo de identificar os fatores presentes nos dados, isto é, conjuntos de características visuais correlacionadas. Essa primeira análise entrega ao pesquisador o gráfico de escarpa ou sedimentação (scree plot) que mede a variação do corpus captada em cada fator, ilustrada na Figura 3.

Figura 3 - Scree plot dos eigenvalues da análise fatorial



fonte: Captura de tela do output fatorial, pelo orientador deste trabalho.

Em uma análise fatorial, o gráfico de sedimentação representa a participação de cada eigenvalue, ou autovalor, na variação encontrada – em ordem decrescente, até que toda variação seja explicada. Assim, como mostrado na Figura 3, os primeiros *eigenvalues* apropriam-se dos maiores percentuais de variação total. A partir de um certo ponto, os eigenvalues não contribuem mais decisivamente para explicar a variação, sendo nove o ponto escolhido para ser o ponto de corte no número de fatores, e assim compor a análise fatorial final.

De posse do número de 9 fatores, uma nova análise fatorial foi realizada, dessa vez a rotacionada, onde o número de fatores, identificado pelo pesquisador no passo anterior, foi utilizado. Tais fatores representam o conjunto de variáveis correlacionadas que nos permite realizar a análise interpretativa, qualitativa. Após essa interpretação, os fatores são elevados à condição de 'dimensões', isto é, de parâmetros subjacentes da variação entre as imagens. O quadro abaixo apresenta uma síntese desses conceitos.

Tabela 2 - Síntese de conceitos envolvidos na análise fatorial e multidimensional

Fatores	Cada fator compreende as variáveis correlacionadas do corpus
Dimensões	Cada dimensão compreende um ou mais discursos presente em cada fator
Polos	Cada polo é um dos componentes do contínuo representado pela dimensão pela dimensão, correspondendo a um discurso predominante

Fonte: desenvolvida pelo pesquisador.

Como parte da análise estatística, foi feito o cômputo de um escore de fator para cada imagem. Esse escore representa a soma, para cada imagem, das características visuais que carregaram em cada polo do fator, subtraindo a soma das características do polo positivo da soma das características do polo negativo. Por exemplo, averiguando que o polo positivo do fator 1 compreenda as características visuais 'combate', 'soldado' e 'guerra', e o polo negativo, 'imagens em preto-e-branco', 'natureza morta' e 'simetria', e que a imagem 1 contenha as características visuais 'guerra' e 'soldado' apenas, todas do polo positivo, o escore da imagem 1 será o resultado da soma das características presentes no polo positivo menos a soma das características presentes no polo negativo.

Uma vez calculados os escores individuais de cada imagem, foram calculados os escores médios de cada nação em cada fator, fazendo a média simples, ou seja, somando os escores de cada imagem que carregou em cada país em cada fator e dividindo pelo número total das fotos que carregaram neste país, neste fator.

Como o foco da pesquisa são as representações do Brasil, as demais nações são usadas apenas para fins de comparação, para colocar em relevo as representações brasileiras, não sendo o objetivo identificar suas próprias representações. O contraste com a representação de um outro país nos presta apenas para colocar em relevo o que o Brasil enfoca.

A fim de detectar se há diferenças estatísticas entre os escores médios de cada nação, foi empregado o teste ANOVA⁹, que comparou, estatisticamente, as médias, revelando se as diferenças notadas são de fato matematicamente válidas. Ao teste F, é associado um valor de significância p , que revela o grau de probabilidade de as diferenças terem ocorrido por acaso. Quando temos um valor de $p < 0,05$, podemos concluir que há uma diferença significativa entre as nações, desse modo podendo considerar que as representações do Brasil são estatisticamente distintas das dos outros países. Caso o valor de p seja superior a 0,05, a conclusão seria a oposta, ou seja, de que não há diferenças estatísticas entre as nações. Além do valor de F, é calculado o R quadrado, por meio do qual é mensurado o tamanho da diferença. Esse valor de R é multiplicado por 100, para que indique a porcentagem de variação pela qual as diferentes nações respondem.

Conforme colocado, a metodologia da Análise Multidimensional Lexical (BERBER SARDINHA, 2014; 2017; 2019; 2020; 2021; BERBER SARDINHA; FITZSIMMONS-DOOLAN, in prep.), é derivada da Análise Multidimensional de cunho funcional (BIBER, 1988; BERBER SARDINHA; VEIRANO PINTO, 2014, 2019), cuja finalidade é apontar conjuntos de itens lexicais relacionados. No caso deste estudo, supõe-se que existem diferentes discursos acerca do que vem a ser “brasilidade”, expressos nas fotos que possibilitam sintetizar a variação lexical em termos de padrões dimensionais (KAUFFMANN, 2020; DELFINO; ARAUJO; BERBER SARDINHA, 2018).

Ainda que a AMD lexical se utilize da mesma técnica estatística da AMD funcional – a análise fatorial que resulta em fatores e, posteriormente, em dimensões –, ela se diferencia por focar nas variáveis lexicais para análise, e não

⁹ *Analysis of variance* (**ANOVA**) é uma fórmula estatística usada para comparar as variâncias entre as medianas (ou médias) de grupos diferentes.

nas de natureza gramatical/funcional, como na Análise Multidimensional preconizada por Biber. Além disso e, principalmente, ao contrário da AMD tradicional, a AMD Lexical permite a identificação de discursos. As resultantes dimensões discursivas da Análise Multidimensional Lexical são, portanto, determinadas pela análise das variáveis lexicais presentes em cada fator.

Conforme relatado anteriormente, a primeira análise fatorial, não rotacionada, produz o *scree plot*. Neste estudo, após a inspeção visual no *scree plot*, optou-se por um ponto de corte no nono fator. Observa-se, nesse momento, uma inflexão na curva que produz um efeito horizontal, a partir do qual os pontos seguintes têm pouco diferencial contributivo para explicar a variação restante. Assim, nove fatores foram a resultante da análise fatorial rotacionada, cujas variáveis serão apresentadas no próximo capítulo.

3. RESULTADOS

A análise fatorial com 388 variáveis lexicais realizada sobre o corpus de estudo resultou nos dados exibidos e interpretados nesta seção. Inicia-se com a apresentação da composição de cada fator resultante da análise fatorial final e subsequente interpretação – incluindo a identificação de semelhanças e diferenças nos fatores obtidas pelo exame dos escores médios das imagens das quatro nações envolvidas e, particularmente, em relação ao Brasil –, com exemplos que ilustram a opção interpretativa imagética da dimensão nomeada do fator, em cada polo. Por fim, segue-se um quadro que resume os achados da análise, útil para a discussão a respeito dos variados discursos em que os conceitos de brasilidade são encontrados nas imagens das redes sociais.

A cada um dos nove fatores deu-se, em seus polos positivos e negativos¹⁰, um sucessivamente, uma nomenclatura titular, curta, como síntese do discurso contido em cada fator, listada abaixo:

Tabela 3 - Títulos curtos de cada variável discursiva

Dimensão	Polo	Discurso
1	Positivo	O Bélico
	Negativo	O Sensível
2	Positivo	Natureza
	Negativo	Paisagem Artificial
3	Positivo	Relevo Topográfico
	Negativo	Bandeiras
4	Positivo	Ambiente Urbano
	Negativo	Pequenos seres
5	Único	Natureza Manipulada e Fértil
6	Positivo	Relíquias de coleção
	Negativo	Esculturas
7	Positivo	Atividades de lazer
	Negativo	Grandes Aparatos
8	Positivo	Tecnologia da Engenharia
	Negativo	Edificações Históricas
9	Positivo	A Vegetação
	Negativo	O registro turístico

Fonte: Elaborado pelo autor.

A partir dessa nomenclatura sintética, a interpretação gradual abaixo trará um detalhamento analítico de cada uma, a fim de explorar o discurso de cada fator numa identificação longa.

3.1 Dimensão 1: Representações da estética do bélico em coexistência com a do sensível

O padrão fatorial do polo positivo deste primeiro fator aparece abaixo, no qual estão apontadas 58 variáveis discursivas:

¹⁰ Em LC, negativo e positivo não tem conotação de juízo de valor ou de descrição literal, mas sim do apontamento estatístico da coexistência e alternância de dois discursos numa mesma dimensão ou recorte textual (BIBER, 1988; BERBER SARDINHA, 2004).

Tabela 4 - Polo positivo do Fator 1

Variável	Carga fatorial
organization	1,19834
military_officer	1,19740
Squad	1,19507
ballistic_vest	1,19463
grenadier	1,18767
gun_barrel	1,18251
camouflage	1,17895
non-commissioned_officer	1,17659
Troop	1,17567
cargo_pants	1,16348
military_camouflage	1,15928
military_rank	1,15581
Marines	1,15434
military_uniform	1,15113
law_enforcement	1,14803
Shotgun	1,14699
machine_gun	1,14441

shooting	1,13661
Army	1,13305
military_organization	1,13034
Military	1,11934
Infantry	1,11698
Combat	1,11674
Soldier	1,11523
military_person	1,11332
Security	1,11198
Swat	1,10169
air_gun	1,08142
army_men	1,07488
Uniform	1,06565
airsoft_gun	1,03452
paintball_equipment	1,03376
Glove	1,00950
Helmet	0,98847
Crew	0,96877
Trigger	0,96144
backpack	0,94190

Boot	0,94155
government_agency	0,90636
Oficial	0,87724
personal_protective_equipment	0,87700
Goggles	0,82507
military_vehicle	0,81633
Belt	0,80439
luggage_and_bags	0,78585
Cap	0,78385
combat_vehicle	0,73345
Bag	0,71411
Games	0,70031
Job	0,69836
Pattern	0,68558
sports_gear	0,67719
armored_car	0,66991
sunglasses	0,64226
Team	0,64137
Hat	0,59404
Recreation	0,57981

Gesture	0,52317
---------	---------

Fonte: Output gerado pelo orientador deste trabalho.

Esse polo concentra características visuais associadas ao poderio bélico institucionalizado, formalizado, de tropas e apresentação formalmente militar, em seus trajes icônicos, portando roupas de camuflagem, equipamento, uniformes, armas e granadas.

No mesmo fator, o padrão fatorial do polo negativo aparece, na Tabela 5, com 26 variáveis:

Tabela 5 - Polo negativo do Fator 1

Variável	Carga fatorial
still_life_photography	-0,58832
Artwork	-0,58521
tints_and_shades	-0,55700
Paint	-0,54236
Line	-0,54128
photography	-0,53756
Grey	-0,53509
symmetry	-0,53300
rectangle	-0,53252
tourist_attraction	-0,52751
Art	-0,52672

black-and-white	-0,48921
Nature	-0,47167
monochrome_photography	-0,46709
Style	-0,45771
illustration	-0,45708
Symbol	-0,45044
Font	-0,44178
electric_blue	-0,44133
visual_arts	-0,43327
stock_photography	-0,41974
Light	-0,37424
monochrome	-0,35931
Darkness	-0,35435
Wood	-0,34961
Blue	-0,32442

Fonte: Output gerado pelo orientador deste trabalho.

Nesse polo negativo, em contrapartida, concentram-se características visuais que remetem à poética e à estética, permeada e manifestada com as linguagens ou técnicas do sensível, como a simetria, a monocromia, o preto-e-branco, o artesanato e a arte.

Tendo em vista que os dois são polos do mesmo fator, a análise mostra que são duas visões antagônicas de representação nacional, dois discursos distintos e opostos:

Discurso 1 - Polo positivo: A imagem nacional está vinculada ao poderio bélico e militar institucionalizado (58 características visuais).

Discurso 2 - Polo Negativo: A imagem nacional está associada às artes visuais, à sensibilidade estética e poética. Nesse campo estão inclusas as linguagens da fotografia, pintura, ilustração e tipografia (26 características visuais).

Como o foco da pesquisa são as representações visuais nacionais do Brasil, foi feita uma comparação estatística por meio de ANOVA, entre os escores das imagens de cada país, conforme mostra a Tabela 6 abaixo.

Tabela 6 - Escores médios de cada país no fator 1

Procedimento GLM – Fator 1

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	-1.5125376	2.7548958
<u>de</u>	984	-2.1402439	2.9205565
<u>gb</u>	948	15.5696203	15.3851780
<u>us</u>	941	-2.1424017	2.9181964

fonte: elaborado pelo autor

Conforme se percebe, o escore médio do Brasil é -1,51, o que significa que as imagens identificadas como representando o Brasil são associadas ao polo negativo do fator, isto é, ao polo que representa o país por meio de uma estética voltada à subjetividade, à sensibilidade, aos sentimentos, ao humano, ao poético. Em contrapartida, a Grã-Bretanha, por sua vez, tem uma imagem ligada ao oposto da imagem predominante do Brasil, ou seja, uma estética da guerra, da força e da destruição. Os demais países da amostra (Estados Unidos e Alemanha) também tiveram sua imagem mais associada ao polo negativo. De modo geral, essa dimensão mostra que há duas estéticas visuais antagônicas nas fotografias do Flickr: uma que evoca o poderio militar, destrutivo, de uma nação, por meio de

fotos de seus soldados em ação, e outra que destaca uma sensibilidade estética para retratar o povo e as cenas do país. O Brasil, portanto, tende a ser caracterizado por meio de uma imagem humanizada, não belicosa, que apela aos sentimentos.

As imagens abaixo ilustram a representação dominante do Brasil no corpus. Note-se que embora a maior variação esteja no polo positivo, do discurso bélico, este é referente à Grã-Bretanha. Nesse fator o Brasil aparece no polo negativo, o da sensibilidade estética.

Figura 4 - Representação típica do Brasil na dimensão 1, polo negativo



Fonte: Amostra do *corpus* Flickr

Figura 5 - Representação do Brasil na dimensão 1, polo negativo



Fonte: Amostra do *corpus* Flickr

Figura 6 - Representação típica do Brasil na dimensão 1, polo negativo



Fonte: Amostra do *corpus* Flickr

Contrastando com essa estética brasileira, apresentamos o exemplo a seguir, que ilustra a estética típica belicosa do polo positivo.

Figura 7 - Representação típica belicosa (Grã-Bretanha) na dimensão 1, polo positivo



Fonte: Amostra do *corpus* Flickr

Diferentemente dessa visão bélica, típica do polo positivo, as imagens sobre o Brasil que pontuaram no polo positivo possuem uma estética peculiar. Enquanto para a Inglaterra a representação foi ligada às corporações militares, os "uniformes" que aparecem no polo positivo brasileiro são, muitas vezes, uniformes de futebol, semelhantes a chamada estética "*brazilcore*". A não ser pelo fator de serem uniformes, em nada tem a ver com os uniformes militares, conforme ilustram as imagens abaixo.

Figura 8 - Representação do Brasil no Fator 1, polo positivo



Fonte: Amostra do *corpus* Flickr

Figura 9 - Representação do Brasil no Fator 1, polo positivo



Fonte: Amostra do *corpus* Flickr

3.2 A beleza natural litorânea frente às paisagens artificiais (Dimensão 2)

O padrão fatorial do polo positivo do segundo fator aparece na Tabela 7, abaixo.

Tabela 7 - Polo positivo do Fator 2

Variável	Carga fatorial
Caribbean	1,13761
Tide	1,11813
Shore	1,10093
coastal_and_oceanic_landforms	1,09373

Bay	1,07222
Tropics	1,05379
wind_wave	1,03085
Inlet	1,02759
Island.	1,02635
Coast	1,01843
people_on_beach	1,01616
Beach	1,00480
Wave	0,99781
Sea	0,97946
Sound	0,95548
Headland	0,94966
Ocean	0,94032
Watercraft	0,91459
Boat	0,89747
Calm	0,89419
water_transportation	0,88241
Lake	0,86942
Loch	0,86607
Water	0,86069

Ship	0,84969
water_resources	0,83931
lake_district	0,81868
Waterway	0,80877
Channel	0,77937
Bank	0,76788
body_of_water	0,74818
Reservoir	0,73673
Summer	0,73515
Horizon	0,71030
Dawn	0,70879
Afterglow	0,69944
River	0,69210
naval_architecture	0,65492
Vacation	0,64439
Watercourse	0,61504
Holiday	0,61331
Wind	0,60823
Reflection	0,59785
palm_tree	0,59237

Sunset	0,59223
Sand	0,58958
Dusk	0,58012
Sunrise	0,57753
Atmosphere	0,49442
Leisure	0,48461
Tourism	0,47512
Evening	0,47456
Azure	0,45487
Cloud	0,43610

Fonte: Output gerado pelo orientador deste trabalho.

No padrão fatorial acima, temos a imagem nacional associada à paisagem costeira e litorânea, ao mar e ao oceano, ao vento e às ondas, associados a férias, verão e lazer. Noutras palavras, um discurso de contemplação e fruição da beleza litorânea.

Em seu polo negativo, temos, conforme ilustrado na Tabela 8:

Tabela 8 - Polo negativo do Fator 2

Variável	Carga fatorial
Relief	-0,68084
Artifact	-0,65024
Statue	-0,63843

Ceiling	-0,63014
Sculpture	-0,62946
Fixture	-0,58891
Door	-0,57849
Brick	-0,55583
Circle	-0,52434
Metal	-0,49600
working_animal	-0,32658

Fonte: Output gerado pelo orientador deste trabalho.

Ou seja, pontuaram como polo negativo imagens de nação associadas a peças artísticas ou de natureza arquitetônica que destacam espaços e passagens, que se traduz como a materialidade e os elementos concretos das paisagens fotográficas, urbanas e sem falseamento dos cenários, porém artificial e construída pelo homem.

Temos então na segunda dimensão mais dois discursos:

Discurso 2 (Polo positivo) – O país representado como beleza natural praiana (54 características visuais);

Discurso 2 (Pólo Negativo) – O país representado como cenário artificial e humano (11 características visuais).

Na análise do teste ANOVA efetuado na comparação das quatro nações incluídas no estudo, conforme exibido na Tabela 9, verificamos que os escores médios obtidos pelo Brasil (5,35) são os mais altos na Dimensão 2, na comparação entre esse país e a amostra das demais nações analisadas. No terreno positivo, mas em menor escala, estão presentes as imagens que representam a Alemanha

(4,49) e os EUA (4,10). Com o escore médio praticamente neutro (0,2) está o Reino Unido. Uma interpretação possível para essa pontuação é que o Brasil representa melhor uma ideia de destinação turística associada ao lazer costeiro, embora tais cenários imagéticos estejam também no imaginário de representação das nações alemã e estadunidense. Conclui-se que o Brasil se posiciona com destaque como paisagem litorânea atrativa associada à ideia de recreação.

Tabela 9 - Escores médios de cada país no fator 2

Procedimento GLM – Fator 2

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	5.35205617	8.92682368
<u>de</u>	984	4.48780488	7.05059853
<u>gb</u>	948	0.20042194	3.22213828
<u>us</u>	941	4.10626993	6.63848972

fonte: elaborado pelo autor

Abaixo, algumas imagens extraídas do corpus ilustram a representação imagética do Brasil no Flickr que obtiveram escores elevados no polo positivo da Dimensão 2:

Figura 10 - Representação típica do Brasil na dimensão 2, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Figura 11 - Representação típica do Brasil na dimensão 2, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Figura 12 - Representação típica do Brasil na dimensão 2, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Avançando com uma imagem para o polo negativo da dimensão 2, abaixo está a Figura 13, selecionada do subgrupo da amostra. Nela é possível notar já a

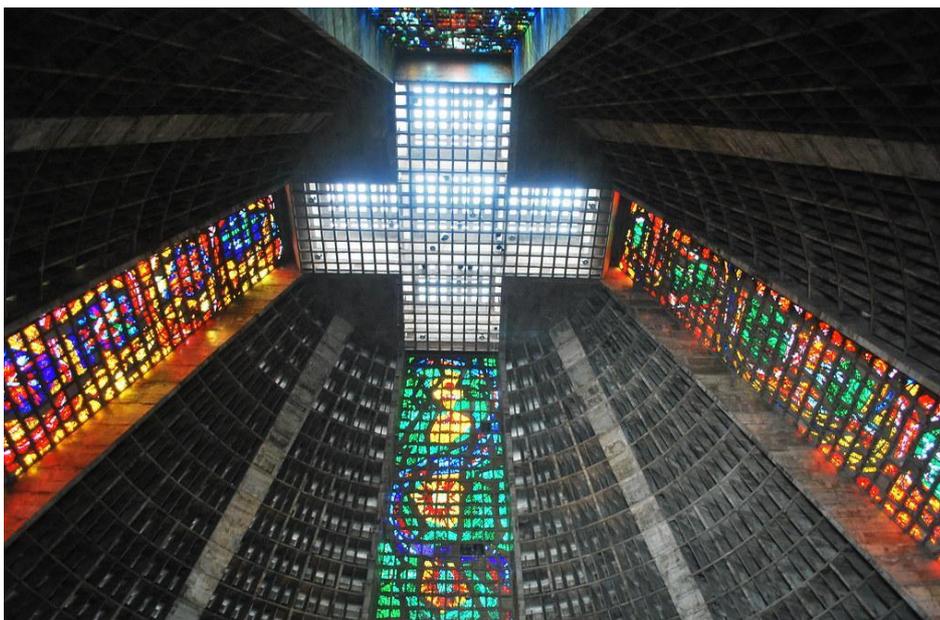
presença contraposta do discurso de cenas artificiais, construções rústicas ou de arquitetura brutalista, colonial e eclética, diferente do polo positivo, predominado por belas paisagens naturais.

Figura 13 - Representação de imagem típica da dimensão 2, polo negativo



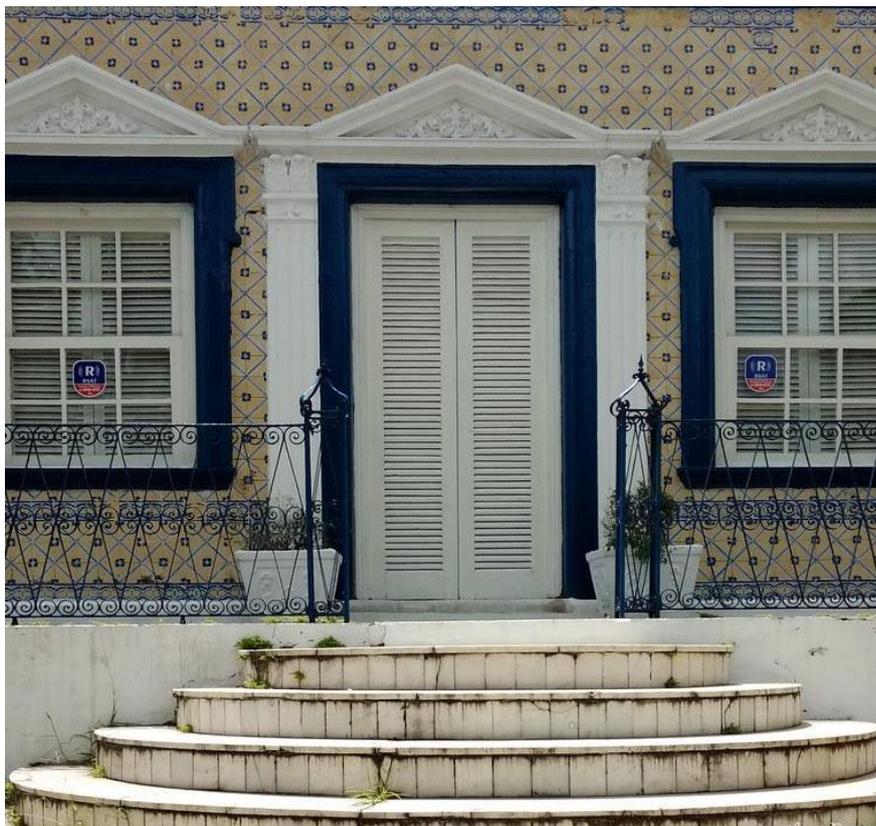
Fonte: Amostra do *corpus* br do Flickr

Figura 14 - Representação de imagem típica da dimensão 2, polo negativo



Fonte: Amostra do *corpus* br do Flickr

Figura 15 - Representação de imagem típica da dimensão 2, polo negativo



Fonte: Amostra do *corpus* br do Flickr

3.3 O registro topográfico e o simbolismo político nacional (Dimensão 3)

O padrão fatorial do polo positivo do terceiro fator aparece na Tabela 10 abaixo. Aqui é notória a variação do discurso para o assunto fotográfico do relevo geográfico ou topográfico.

Tabela 10 - Variação do discurso no polo positivo do fator 3

Variável	Carga fatorial
intrusion	1,25467
fault	1,23825
makhtesh	1,22276

wadi	1,17285
butte	1,15974
badlands	1,15504
escarpment	1,13152
erosion	1,12610
plateau	1,07010
formation	1,07002
outcrop	1,06435
bedrock	1,01768
cliff	1,00896
ridge	1,00567
valley	0,99526
terrain	0,98815
summit	0,98714
chaparral	0,96945
klippe	0,96907
Massif	0,92623
geological_phenomenon	0,91576
geology	0,90870
mountainous_landforms	0,89351

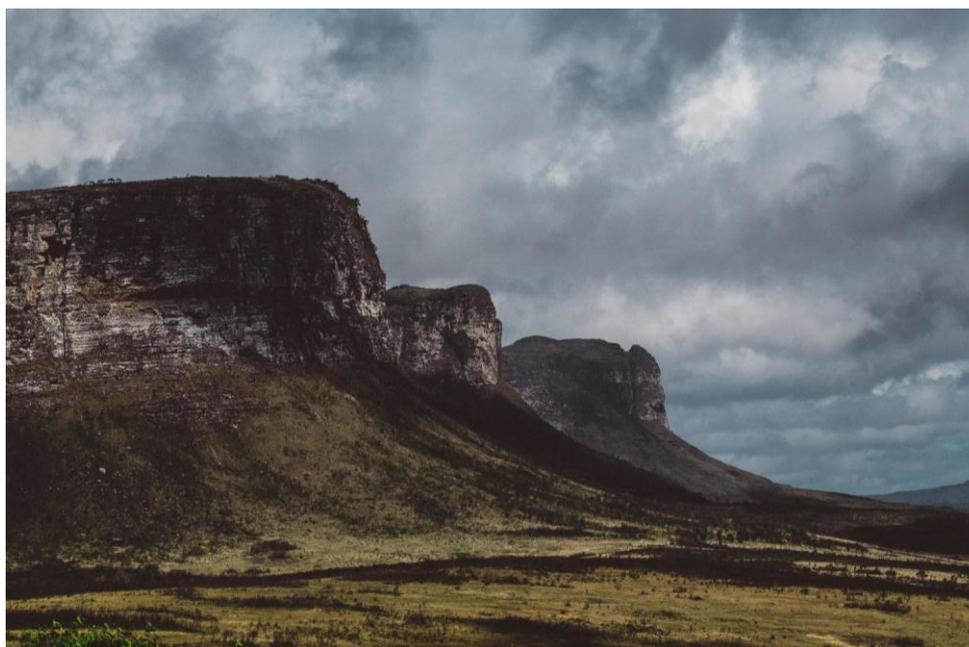
mountain	0,87830
shrubland	0,83949
aeolian_landform	0,82698
Fell	0,79904
mountain_range	0,76895
Hill	0,76824
Rock	0,76681
highland	0,74289
Tundra	0,72803
Trail	0,67836
plant_community	0,66796
aerial_photography	0,62693
landscape	0,59660
Arch	0,58165
historic_site	0,54304
adventure	0,52496
wilderness	0,52066
Soil	0,49735

Fonte: Output gerado no SAS pelo orientador deste trabalho.

As 41 variáveis acima apresentam o polo positivo da Dimensão 3, com imagens nacionais voltadas para o relevo topográfico, com montanhas, que

compõem paisagens ao lado de penhascos, vales e despenhadeiros. Amplos horizontes e espaços luminosos com céus e nuvens complementam as imagens desse polo.

Figura 16 - Representação típica do Brasil na dimensão 3, polo positivo



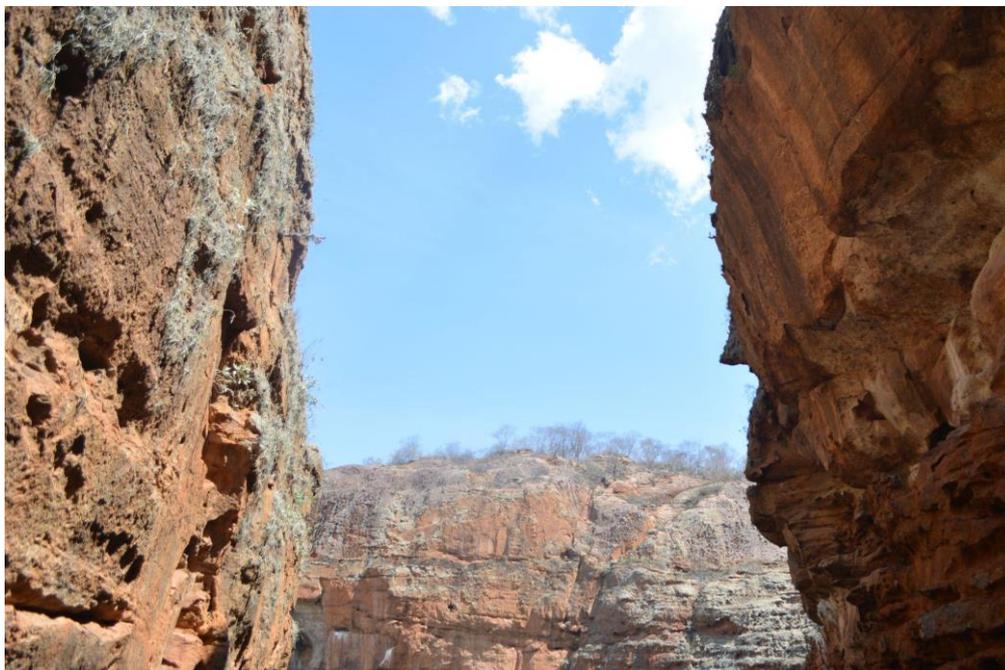
Fonte: Amostra do *corpus* br do Flickr

Figura 17 - Representação de Brasil no fator 3, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Figura 18 - Representação típica do Brasil na dimensão 3, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Figura 19 - Representação típica do Brasil na dimensão 3, polo positivo



Fonte: Amostra do *corpus* br do Flickr

Já em contrapartida, o polo negativo da Dimensão 3 apresenta, em sua maioria, imagens associadas a mastros e bandeiras, inscritas em eventos e salas, por vezes na forma de desenhos. Isto é, há presença daqueles símbolos relacionados ao interesse nacionalista ou ao estudo das bandeiras, a vexilologia. Além das bandeiras, há ainda outras imagens relacionadas a objetos de decoração. A lista de 7 variáveis lexicais do polo negativo da Dimensão 3 é apresentada na Tabela 11, abaixo.

Tabela 11 - Fator 3, carga fatorial do polo negativo

Variável	Carga fatorial
Drawing	-0,50283
Room	-0,49374
Glass	-0,47778
Window	-0,47412
Flag	-0,42866
Pole	-0,42840
Event	-0,38946

Fonte: Output do SAS, processado pelo professor orientador.

A seguir, vemos algumas imagens que carregaram nesse polo:

Figura 20 - Representação de Brasil, dimensão 3, polo negativo



Fonte: Corpus br do Flickr

Figura 21 - Representação típica do Brasil na dimensão 3, polo negativo



Fonte: Corpus br do Flickr

Figura 22 - Representação típica do Brasil na dimensão 3, polo negativo



Fonte: Corpus br do Flickr

3.4 O ambiente urbano e os pequenos seres (Dimensão 4)

O padrão fatorial do polo positivo da quarta dimensão aparece na Tabela 12 abaixo. Nota-se aqui marcações discursivas referenciadas ao ambiente urbano:

Tabela 12 - Fator 4 e suas variáveis do polo positivo

Variável	Carga fatorial
----------	----------------

Metropolis	1,04992
mixed-use	1,02293
Downtown	0,97924
commercial_building	0,95874
Apartment	0,94852
neighbourhood	0,94811
Cityscape	0,93694
town_square	0,93304
condominium	0,91680
tower_block	0,90368
Skyline	0,89894
Suburb	0,89308
urban_design	0,87929
City	0,87903
urban_area	0,87753
Skyscraper	0,85176
medieval_architecture	0,84037
Spire	0,82266
Building	0,80518
Facade	0,79047

infrastructure	0,79030
Tower	0,78176
architecture	0,73481
Church	0,72694
Roof	0,70928
Landmark	0,70575
Chimney	0,69527
Brickwork	0,69501
House	0,68413
classical_architecture	0,68197
Electricity	0,68114
Town	0,66777
Dome	0,66393
Bridge	0,65336
Daytime	0,64256
place_of_worship	0,63101
Pedestrian	0,60396
Canal	0,59926
Daylighting	0,58321
Village	0,57356

Street	0,57298
sash_window	0,53965
Column	0,53645
Night	0,51683
World	0,49728
Sky	0,46050
Morning	0,41995

Fonte: Output do SAS, processado pelo professor orientador.

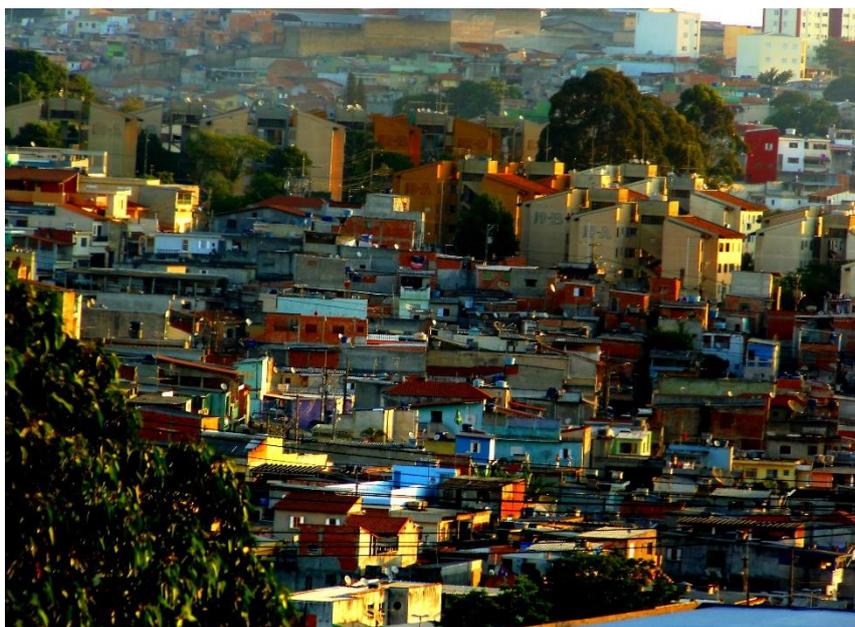
O polo positivo da dimensão quatro é caracterizado por cenas urbanas em que convivem fachadas de edifícios clássicos e modernos, criando espaços disciplinados de metrópoles, em 47 variáveis. Centros urbanos alternam espigões e outros equipamentos viários e criam horizontalmente uma paisagem urbanizada, porém salpicada com subúrbios e espaços vacantes. A representação imagética que o Brasil expressa nesse polo é a da realidade urbana do país, na qual convivem elementos arquitetônicos tradicionais em meio aos prédios urbanos das cidades brasileiras, incluindo seus núcleos formados em bairros periféricos.

Figura 23 - Cena representando o Brasil no polo positivo do fator 4



Fonte: Corpus br do Flickr

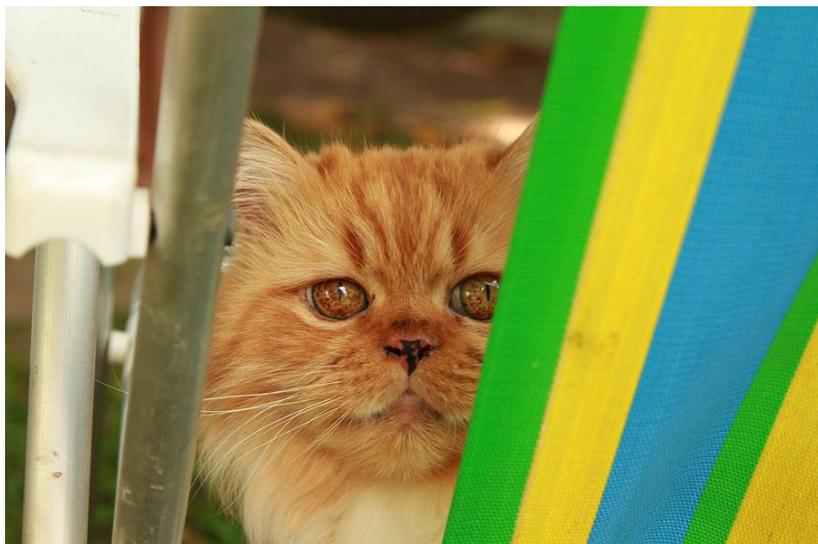
Figura 24 - Cena urbana representando o Brasil. Fator 4, polo positivo.



Fonte: Corpus br do Flickr

Outro discurso emergente do quarto fator são os que têm escores negativos, conforme mostrado na Tabela 13, abaixo das imagens desse polo. As representações associadas ao polo negativo apresentam imagens de animais de superfície com focinhos, pêlos e bigodes, predominantemente pequenos. A fauna silvestre complementa essa representação imagética, com 11 variáveis nesse polo negativo.

Figura 25 - Imagem de representação do Brasil, no discurso do polo negativo, fator 4



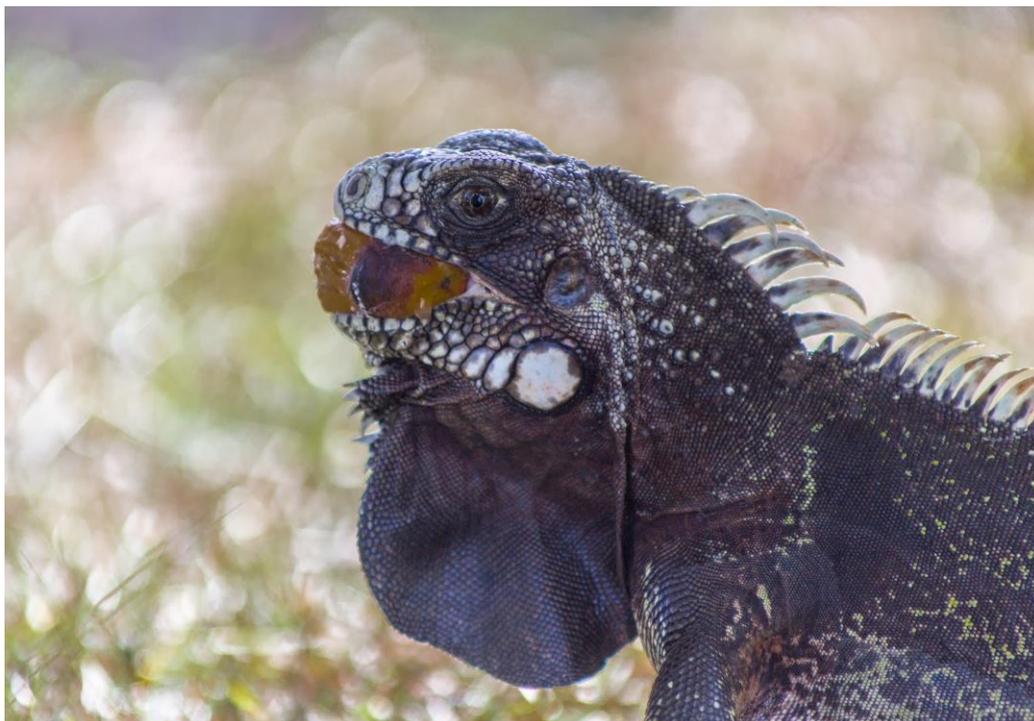
Fonte: Corpus br do Flickr

Figura 26 - Imagem de animais representando o Brasil no polo negativo do quarto fator



Fonte: Corpus br do Flickr

Figura 27 - Imagem representando o Brasil no polo negativo do quarto fator



Fonte: Corpus br do Flickr

Figura 28 - Representação de Brasil no polo negativo do fator 4



Fonte: Corpus br do Flickr

Figura 29 - Representação de Brasil no polo negativo do fator 4



Fonte: Corpus br do Flickr

Tabela 13 - Polo negativo do quarto fator

Variável	Carga fatorial
Eye	-0,78928
Whiskers	-0,78917
terrestrial_animal	-0,72202
Head	-0,72085
Felidae	-0,71862
Chest	-0,70937
Snout	-0,65582

Carnivore	-0,62291
Fur	-0,55614
fictional_character	-0,52485
Wildlife	-0,51447

Fonte: Processamento SAS, output pelo orientador da pesquisa.

Na análise para detectar possíveis diferenças entre representações nacionais e, simultaneamente, salientar o posicionamento da representação de Brasil que o distingue de outras culturas ocidentais, o país é percebido no polo urbano, no terreno positivo do fator 4, como observado na Tabela 14. Os escores de representações nacionais de referência também transitam no polo positivo, liderados pelos grupos Alemanha (4,74) e EUA (2,98), tendo o Reino Unido valor inferior (1,60).

Tabela 14 - Escores médios de cada país no fator 4

Procedimento GLM – Fator 4

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
br	997	2.68004012	4.67400450
de	984	4.77439024	5.72655241
gb	948	1.60021097	3.85282024
us	941	2.98299681	5.47972355

fonte: elaborado pelo autor, com output SAS.

3.5 A natureza manipulada e fértil (Dimensão 5)

O padrão fatorial do polo positivo, o único polo do quinto fator, aparece na Tabela 15 abaixo.

Tabela 15 - Fator 5, com um único polo

Variável	Carga fatorial
----------	----------------

Farm	1,20865
Agriculture	1,08868
Plantation	1,05525
Field	1,04620
grass_family	1,00178
Prairie	1,00018
land_lot	0,99148
Pasture	0,98250
Meadow	0,97175
Steppe	0,96123
Plain	0,95116
rural_area	0,83345
Grassland	0,82521
Wildflower	0,81427
dirt_road	0,75224
flowering_plant	0,74474
Grass	0,73935
Ecoregion	0,71701
Landscaping	0,71514
people_in_nature	0,70335

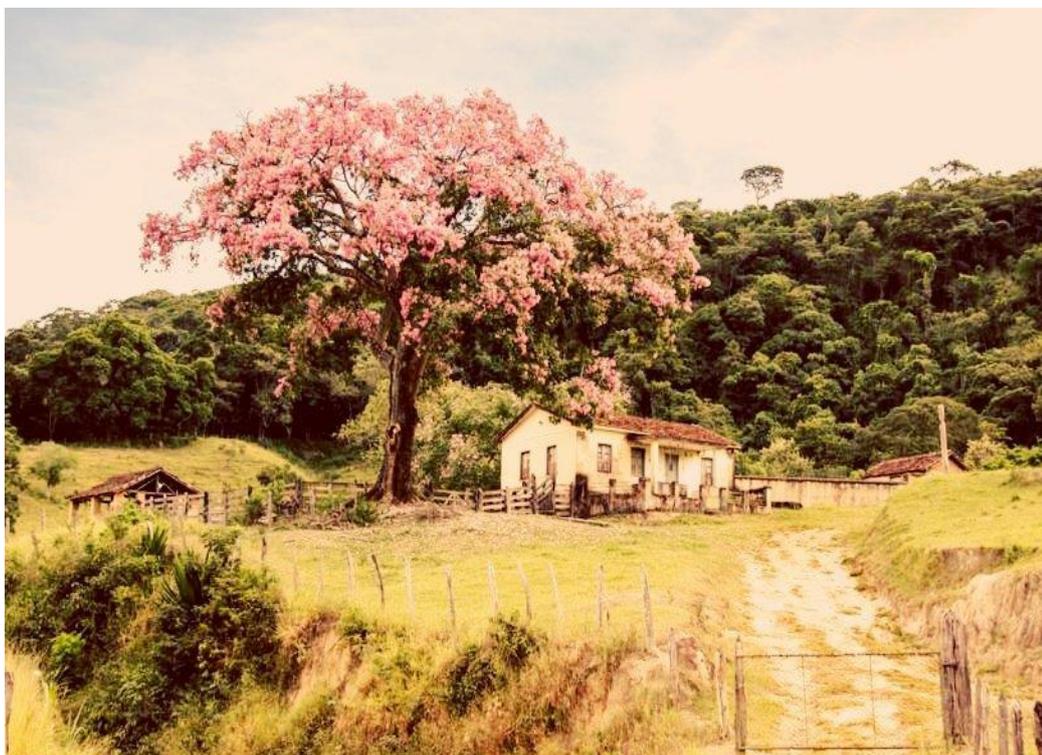
plant_stem	0,68833
Vegetation	0,66725
natural_landscape	0,65107
Plant	0,62590
natural_environment	0,61185
Shrub	0,60916
Flower	0,58278
Cottage	0,56637
terrestrial_plant	0,54721
Sunlight	0,54135
Garden	0,52507
meteorological_phenomenon	0,51956
Slope	0,51127
Shade	0,50620
Cumulus	0,49549
Shadow	0,45179
Tree	0,43549
atmospheric_phenomenon	0,41556

Fonte: Output do processamento SAS

O fator 5 se distingue porque nele não coexistem dois discursos, o que não significa qualquer anomalia a ser destacada. Aqui, um polo único imagético, com

38 variações, que concentra representações nacionais caracterizadas por paisagens agrícolas compostas por fazendas, plantações, pastos e campos rurais, aliando ambiente natural com vegetação e paisagismo.

Figura 30 - Representação de Brasil no fator 5



Fonte: Corpus br do Flickr

Figura 31 - Representação de Brasil no Fator 5



Fonte: Corpus br do Flickr

A análise gerada pela comparação das médias nacionais brasileira e de referência na Dimensão 5 indicam um uso moderado na referência que remete visualmente a fazendas e campos com plantações, mais utilizado por referenciais imagéticos associados à Alemanha (4,95) e aos EUA (3,84). A média do Brasil no polo é de 2,60, apenas superior às representações associadas ao Reino Unido (1,80).

Tabela 16 - Escores de cada país no quinto fator**Procedimento GLM – Fator 5**

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	2.59679037	3.45334797
<u>de</u>	984	4.99491870	6.06047638
<u>gb</u>	948	1.79535865	2.90328067
<u>us</u>	941	3.83953241	3.92547184

fonte: elaborado pelo autor

3.6 Relíquias de valor simbólico, icônico ou místico (Dimensão 6)

Expondo agora o padrão fatorial do polo positivo do sexto fator, na Tabela 17 abaixo, temos um discurso materializado em 25 variáveis.

Tabela 17 - Fator 6, variáveis do polo positivo

Variável	Carga fatorial
sports_car	1,19109
antique_car	1,12860
compact_car	1,06538
Rim	1,06391
vintage_car	1,06252
Racing	1,02538
notchback	0,99277
luxury_vehicle	0,99257
Car	0,99124
motorsport	0,98421

Parking	0,93411
Vehicle	0,92813
off-road_vehicle	0,91484
Tread	0,90711
Rolling	0,86272
Truck	0,81567
convertible	0,78853
Traffic	0,78420
mode_of_transport	0,76608
Asphalt	0,69553
transport	0,68431
motorcycle	0,66937
Lane	0,64427
Road	0,56746
road_surface	0,48961

Fonte: Output do processamento fatorial, no SAS.

O fator 6 concentra em seus polos discursos distintos também, associados às imagens nacionais: no polo positivo estão variáveis lexicais ligadas a veículos automotores, sejam esportivos, de luxo, para colecionadores, de formato compacto ou conversível, além de motocicletas e elementos viários. Escolheu-se nomear essa variável suscintamente como relíquias de valor simbólico.

Figura 32 - Imagem representando o Brasil no polo positivo do fator 6



Fonte: Corpus br do Flickr

Figura 33 - Representação de Brasil, polo positivo do fator 6



Fonte: Corpus br do Flickr

Concomitante ao discurso acima e conforme apresentado na Tabela 18, a seguir, o polo negativo do fator 6 apresenta imagens associadas às variáveis lexicais descritivas de estatuária em pedra, esculturas clássicas, mitologia e monumentos.

Figura 34 - Representação de Brasil, polo negativo do fator 6



Fonte: Corpus br do Flickr

Figura 35 - Representação de imagem típica do Brasil na dimensão 6, polo negativo



Fonte: Corpus br do Flickr

Figura 36 - Imagem de representação do Brasil, polo negativo do fator 6



Fonte: Corpus br do Flickr

As imagens expressam o discurso que aparecem assim listado:

Tabela 18 - Polo negativo do fator 6

Variável	Carga fatorial
stone_carving	-0,77810
classical_sculpture	-0,71610
carving	-0,69969
mythology	-0,66359
monument	-0,62679

Fonte: Output obtido pelo orientador da pesquisa.

O Brasil está mais claramente associado à série de imagens representadas pelo polo positivo da Dimensão 6, com escore médio de 2,36 (cf. Tabela 19). Alemanha (1,35), EUA (0,95) e Reino Unido (0,81) estão posicionados com escores discretamente inferiores. Esse resultado é significativo estatisticamente – ou seja,

ele sugere que a representação de imagens de veículos sensibiliza mais quem quer associá-las ao Brasil, o que indica uma influência cultural eminentemente urbana; No entanto, pode-se notar que não tais carros exprimem um simbolismo de relíquias, típico do colecionismo automobilístico, podendo significar também um saudosismo, de contexto político-econômico ou da prática de hobby, em fotografias poéticas.

Tabela 19 - Escores da dimensão 6, por país

Procedimento GLM – Fator 6

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	2.36008024	3.58456799
<u>de</u>	984	1.34654472	2.50387863
<u>gb</u>	948	0.81012658	2.47859320
<u>us</u>	941	0.94580234	2.40925430

fonte: elaborado pelo autor

3.7 A estética em atividades de lazer e aparatos de grande escala (Dimensão 7)

O padrão fatorial do polo positivo do sétimo fator aponta os seguintes discursos:

Tabela 20 - Variação discursiva no polo positivo do Fator 7

Variável	Carga fatorial
Walking	0,84363
t-shirt	0,84294
Crowd	0,81748
Fun	0,80559
Thigh	0,80162

Shorts	0,78628
public_event	0,76895
human_leg	0,75382
Shoe	0,73059
sport_venue	0,71777
eyewear	0,71011
Happy	0,70840
Sports	0,70588
advertising	0,66285
Sitting	0,64314
headgear	0,63306
fashion_accessory	0,59345
Logo	0,58038
graphics	0,57483
Jacket	0,57236
magenta	0,54544
photograph	0,50090
flag_of_the_united_states	0,44203
flooring	0,42343

Fonte: Processamento desenvolvido pelo orientador da pesquisa.

O fator 7 contém dois polos discursivos. No polo positivo do fator, as variáveis lexicais associadas às imagens nacionais destacam elementos como pessoas que caminham ou estão presentes em eventos públicos, que se divertem e vestem camisetas, óculos de sol, bermudas e acessórios. Lembra-nos o movimento do *brazilcore*¹¹ na medida em que é uma estética registrada nessa dimensão em momentos de lazer ou esporte.

Figura 37 - Imagem de Brasil no polo positivo do fator 7



Fonte: Corpus br do Flickr

¹¹ *Brazilcore*, ou *Brazil Aesthetics*, é um movimento fashionista surgido em 2022. A designer de moda Fabrícia Santos apresentou a tendência como resposta ao desafio de desvincular as cores da bandeira do país de movimentos políticos. Santos, Fabrícia, em entrevista. Ver em <https://blog.stz.com.br/moda/brazilcore>

Tabela 21 - Polo negativo do fator 7

Variável	Carga fatorial
gun_turret	-0,82913
Tank	-0,76499
self-propelled_artillery	-0,70352
estate	-0,59335

Fonte: Output do SAS, processado pelo orientador.

Nas imagens dos outros países as imagens ficam mais coerentes com a dimensão, porém na pasta do corpus referente ao Brasil a representação fotográfica referêcia grandes estruturas e grandes propriedades:

Figura 40 - Representação de Brasil no polo negativo do fator 7

Fonte: Corpus br do Flickr

Figura 41 - Representação de Brasil no polo negativo do fator 7



Fonte: Corpus br do Flickr

Figura 42 - Representação de Brasil no polo negativo do fator 7



Fonte: Corpus br do Flickr

Na análise do teste ANOVA, que aponta diferenças entre representações nacionais, houve variação significativa, que indica que a representação nacional do Brasil (média de 1,32) guarda um traço mais pronunciado no polo positivo da Dimensão 7, frente às demais nacionalidades de referência – EUA, 0,73; Alemanha, 0,53; e Reino Unido, 0,44.

Tabela 22 - Escores médios de cada país no fator 7

Procedimento GLM – Fator 7

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
br	997	1.31594784	1.98727792
de	984	0.52845528	1.37895948
gb	948	0.44198312	1.23902109
us	941	0.73113709	1.48071923

fonte: elaborado pelo autor

3.8 A tecnologia da engenharia e das edificações (Dimensão 8)

As variáveis discursivas do fator 8 apresentaram a seguinte composição:

Tabela 23 - Discursos presentes no polo positivo do fator 8

Variável	Carga fatorial
aerospace_manufacturer	1,29155
Flight	1,28810
Aviation	1,26103
air_travel	1,25831
airplane	1,25013
Airline	1,22878

aircraft	1,18734
aerospace_engineering	1,17851
service	0,86818
Engineering	0,70559
Steel	0,55688
public_utility	0,54217
Gas	0,51927
Machine	0,48620

Fonte: Elaborado pelo autor, com output do SAS

O polo positivo do fator 8 relaciona-se com as grandes criações da engenharia, indústria aeroespacial, notadamente a aviação e serviços ligados à aeronáutica.

Figura 43 - Discurso presente no fator 8, polo positivo.



Fonte: Corpus Flickr

Figura 44 - Discurso presente no fator 8, polo positivo



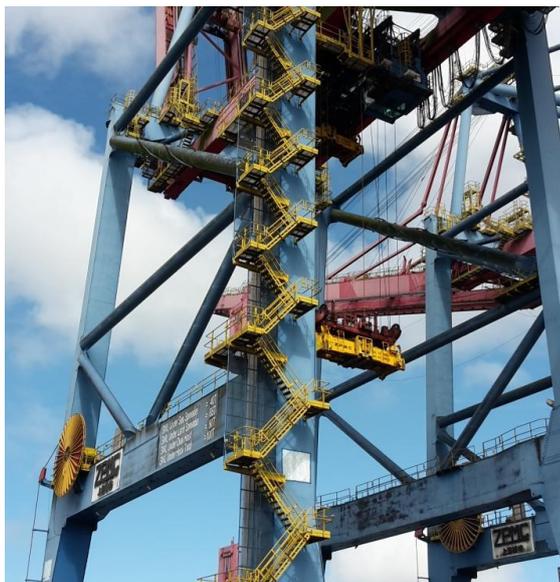
Fonte: Corpus Flickr

Figura 45 - Representação de Brasil, Fator 8 - polo positivo



Fonte: Corpus Flickr

Figura 46 - Representação de Brasil, Fator 8 - polo positivo



Fonte: Corpus Flickr

Figura 47 - Representação de Brasil, Fator 8, polo positivo.



Fonte: Corpus Flickr

Já o polo negativo do fator 8 está ligado a templos e edificações históricas:

Tabela 24 - Discursos presentes no oitavo fator negativo

Variável	Carga fatorial
----------	----------------

trunk	-0,57916
temple	-0,51216
history	-0,42370

Fonte: Output do processamento fatorial

Figura 48 - Representação de Brasil, polo negativo do fator 8



Fonte: Corpus Flickr

Figura 49 - Representação de Brasil no polo negativo do oitavo fator



Fonte: Corpus Flickr

Figura 50 - Representação de Brasil no polo negativo do fator 8



Fonte: Corpus Flickr

Figura 51 - Representações de Brasil, polo negativo no Fator 8

Fonte: Corpus Flickr

A análise comparativa entre as imagens de representação nacionais indica pouca diferenciação entre as diferentes nacionalidades, embora o valor de $p < 0,0001$ indique significância estatística. Todos os escores médios aproximam-se, em maior ou menor medida, de zero. Assim, as representações imagéticas associadas ao Reino Unido, no polo positivo da Dimensão 8, têm a média de 0,35, enquanto os demais grupos nacionais obtiveram escores residuais semelhantes (EUA, 0,08; Alemanha, 0,08; e Brasil, 0,06).

Tabela 25 - Escores médios de cada país no fator 8**Procedimento GLM – Fator 8**

<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	0.06018054	0.80932639
<u>de</u>	984	0.07621951	1.05941669
<u>gb</u>	948	0.35126582	1.63192428
<u>us</u>	941	0.08076514	1.09828776

fonte: elaborado pelo autor

3.9 A vegetação e o olhar turístico-fotográfico (Dimensão 9).

O padrão fatorial do polo positivo do nono fator aparece na Tabela 26 abaixo.

Tabela 26 - Discursos presentes no polo positivo do fator 9

Variável	Carga fatorial
Fir	1,14868
Pine	1,02292
pine_family	0,96805
temperate_broadleaf_and_mixed_forest	0,88777
conifer	0,88747
deciduous	0,80645
tropical_and_subtropical_coniferous_forests	0,79869
woodland	0,75722
Snow	0,72885
freezing	0,70913
branch	0,67511
autumn	0,64986
forest	0,61858
jungle	0,57876
Twig	0,50733
biome	0,49669
winter	0,45981
painting	0,43599

Fonte: Output do processamento em SAS

O polo positivo do fator 9 está associado a uma representação de florestas coníferas temperadas, com ênfase nas épocas de outono e inverno, a presença de neve e outros elementos tais como folhas, galhos, lenha, compondo uma imagem de extremos negativos de temperatura com árvores do hemisfério Norte, havendo, portanto, poucos pontos de contato com uma imagem do Brasil com esse conjunto de variáveis lexicais descritivas do universo imagético associado a nacionalidades. Embora esta dimensão não descreva a realidade da imagem do Brasil, a existência de tal perspectiva é justificada pela recorte do desenho do corpus, escolhido para esta análise para refletir também a ocorrência de outros traços culturais existentes nas culturas alemã, norte-americana e britânica (expoentes de uma cultura dita ocidental); além disso, no recorte brasileiro deste fator, as imagens são árvores coníferas típicas do sul e sudeste brasileiro, como a araucária e o angico (Figura 52 adiante).

Figura 52 - Representação de Brasil, polo positivo do fator 9



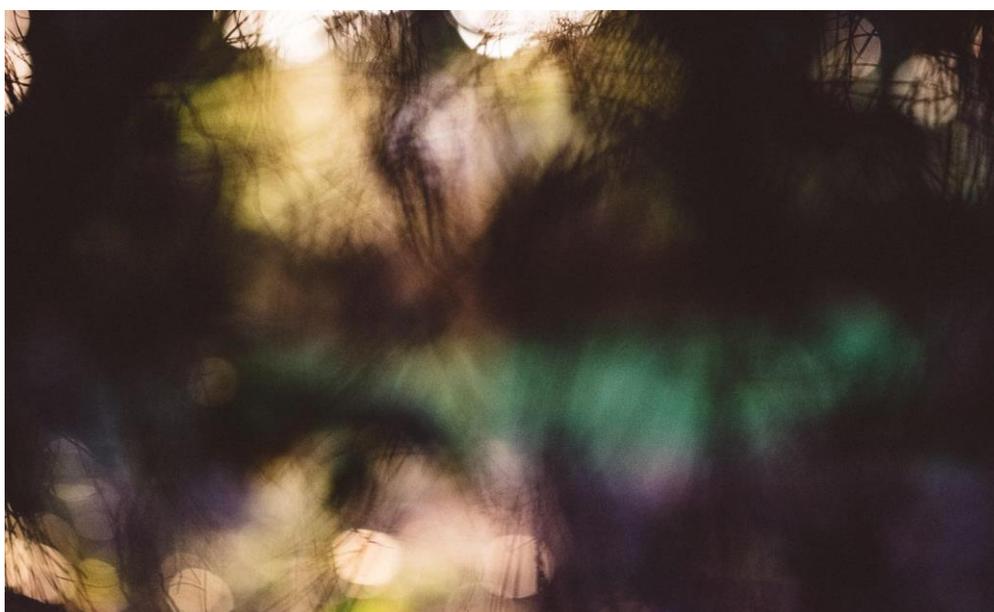
Fonte: Corpus Flickr

Figura 53 - Representação de Brasil, polo positivo do fator 9



Fonte: Corpus Flickr

Figura 54 - Representação de Brasil, polo positivo do fator 9



Fonte: Corpus Flickr

Figura 55 - Representação de Brasil, polo positivo do fator 9



Fonte: Corpus Flickr

Figura 56 - Representação de Brasil, polo positivo do fator 9



Fonte: Corpus Flickr

Figura 57 - Representação de Brasil, polo positivo do fator 9



Fonte: Corpus Flickr

O polo negativo, em relação a outros países do fator 9, contém apenas as variáveis lexicais ligadas à história antiga e ao turismo, o que limita bastante a interpretação de imagens nacionais ligadas a esse discurso. No entanto, as do corpus brasileiro contém imagens típicas do olhar turístico-fotográfico brasileiro.

Figura 58 - Representação de Brasil, polo negativo do fator 9



Fonte: Corpus Flickr

Figura 59 - Representação de Brasil, polo negativo do fator 9



Fonte: Corpus Flickr

Figura 60 - Representação de Brasil, polo negativo do fator 9



Fonte: Corpus Flickr

Figura 61 - Representação de Brasil, polo negativo do fator 9



Fonte: Corpus Flickr

Figura 62 - Representação de Brasil, polo negativo do fator 9

Fonte: Corpus Flickr

Os discursos presentes no polo negativo se alternam com os do polo positivo com apenas duas variações:

Tabela 27 - Variáveis discursivas do polo negativo do fator 9

Variável	Carga fatorial
ancient_history	-0,63835
Travel	-0,49986

Fonte: Output do SAS, processado pelo orientador.

Figura 63 - Imagem do polo negativo, no corpus da Alemanha, do fator 9



Fonte: Corpus Flickr

Apesar de haver significância estatística, é pequena a influência do fator 9 na diferenciação entre países, os EUA representam o maior escore médio entre os quatro subgrupos nacionais – 1,22 –, seguidos por Alemanha (1,00). Com escores mais próximos de zero, estão Brasil (0,28) e Reino Unido (0,20). Ou seja, a Dimensão 9 br pouco contribui na caracterização da imagem nacional brasileira.

Tabela 28 - Escores médios de cada país no fator 9

Procedimento GLM – Fator 9

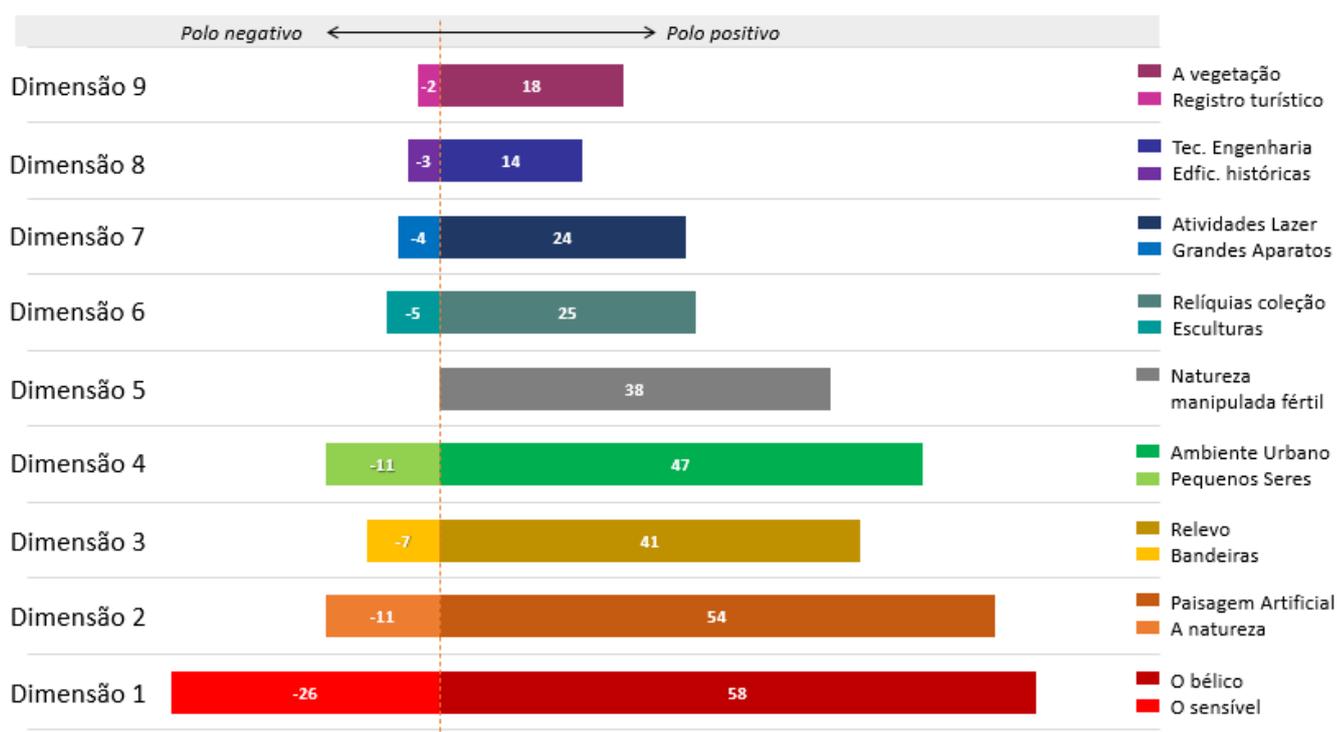
<u>Nível do país</u>	<u>N</u>	<u>Média</u>	<u>Desvio Padrão</u>
<u>br</u>	997	0.27883651	1.33552159
<u>de</u>	984	1.00609756	2.37697574
<u>gb</u>	948	0.20042194	1.03370844
<u>us</u>	941	1.22316684	2.574655861

fonte: elaborado pelo autor

4. DISCUSSÃO

O estudo analisou o corpus imagético obtido do site Flickr (“*corpus Flickr*”). A AMD Lexical conduzida no corpus revelou os discursos que tangenciam empiricamente o conceito de brasilidade por meio das imagens, como visto nas seções anteriores, o que abarca sentidos diversos aos tradicionalmente considerados, quando a fonte para definir tal conceito é composta pelos elementos visuais que integram as redes sociais, que vêm desenvolvendo continuamente novas formas de comunicação integradas e multimodais. Buscando aqui uma discussão dos resultados apresentados acima, para fins de visualização panorâmica, têm-se a seguinte distribuição discursiva, por carga variável, presente nos 9 fatores recortados para este estudo:

Figura 64 - Gráfico piramidal com a variação e difusão discursiva no corpus Flickr



Fonte: Elaborado pelo autor.

Assim, através da análise do corpus imagético representado acima, que buscou as principais representações visuais associadas a quatro nacionalidades e delas isolando-se os discursos ligados ao Brasil, foi possível compor um rico e

sugestivo mosaico em que interagem diversos discursos emergentes e, simultaneamente, se permite traçar um perfil particular do Brasil, diante de outras culturas, como as dos EUA, Reino Unido, Inglaterra e Alemanha. Cada fator é apresentado abaixo, na perspectiva de estabelecer uma descrição dessas características imagéticas do país e discutindo os elementos discursivos manifestados no capítulo anterior.

No fator 1, o Brasil aparece caracterizado como um país ligado a representações artísticas, dotado de sensibilidade estética e poética, longe de uma imagem bélica com instituições militares presentes, que ocorre no polo oposto do mesmo fator. O Brasil possui também uma imagem fortemente associada ao polo positivo do fator 2, que enxerga o país como dotado de uma paisagem com belezas naturais costeiras, em contraposição a uma visão de espaços urbanos que se aproxima do artificial. No fator 3, o Brasil é representado por um território composto por cânions, montanhas e outros acidentes geográficos de interesse geológico, sendo este um discurso mais presente que o da exibição de bandeiras hasteadas em eventos e similares imagens. O fator 4 revela a propensão de o Brasil estar representado por imagens que ligam a paisagem urbana com o patrimônio histórico, superando uma visão mais ligada à exibição de animais da fauna silvestre ou de estimação. A visão que liga o Brasil ao agronegócio surge destacado no fator 5, quando aparecem referências visuais do campo com produção agrícola e o ambiente rural. Como qualquer outra nação ocidental, o Brasil tem aspiração por veículos que propiciem status, como transparece nas imagens que caracterizam o fator 6, enfatizando as imagens de carros esportivos e de luxo – em contraposição a uma ideia de associação com estatuária e monumentos. Já uma associação visual do Brasil por eventos públicos abertos, a estética “*brazilcore*” e torcidas uniformizadas sobressai no fator 7, além de mostrar que o país está distante no polo discursivo que expressa imagens de tanques de guerra e artilharia – fazendo eco ao afastamento do discurso militar ocorrido no polo positivo do fator 1. Por fim, o Brasil é pouco associado aos discursos que são o foco do fator 8, em que se alternam um polo associado a aeronaves e aos grandes feitos da engenharia, como a aviação, e outro, mais ligado a templos históricos. A mesma situação de participação tímida ocorre com a representação do Brasil em relação aos discursos

emergentes no fator 9, interpretados como paisagem de bosques temperados e, de outra parte, o registro fotográfico típico de contextos do turismo.

Assim, os discursos que emergem dos fatores analisados podem servir de fonte para a caracterização nacional do terreno da representação imagética. Reforçado pela presença de escores médios pontuados por nacionalidade e o poder de identificação entre nações em cada fator, identificados pelos testes de ANOVA, a imagem sobre o Brasil também pode ser matizada segundo essa lente de interpretação.

Desse modo, o traço discursivo imagético principal do Brasil é o que revela as belezas litorâneas naturais do país, já que é aquele com maior escore entre os países analisados, no fator 2. Em segundo plano, mas ainda com importância, estão as representações imagéticas expressas pelos fatores 7 e 6, respectivamente os discursos sobre eventos públicos de lazer e os de relíquias simbólicas, fundantes para o imaginário discursivo brasileiro da contemporaneidade, aliando a afirmação de grupo organizado à liberdade individual, na esfera do consumo de bens duráveis de locomoção.

Outros traços definidores da representação imagética do Brasil são os discursos construídos sobre as artes visuais, com ênfase na sensibilidade estética e poética, e aqueles que retratam a paisagem urbana e o patrimônio histórico. Podemos definir que existe em tais enfoques um olhar estético mais cuidadoso, porém situado em um contexto urbano que é resultado de uma revisão e evolução histórica. Os resquícios do passado rural brasileiro residem, por final, esmaecidos para dar lugar aos conteúdos visuais mais modernizados dos campos domesticados da produção agropecuária, assim como o restante de biodiversidade e riqueza florestal foi de certa forma reduzido a imagens ligadas a vales, montanhas e pontos de interesse geológico. Trata-se de um apagamento da biodiversidade, como visto, assim como há um apagamento das referências a conteúdos tecnológicos de ponta, como visto em relação à Dimensão 8.

Em síntese, a representação do Brasil nas redes sociais e o conceito de brasilidade que emerge é aquele ligado a um clichê tropical, que guarda alguma característica

estética e poética na comunicação visual, mas que se rende à sociabilidade e à existência de teor consumista que se apoia nos bens proporcionados pela vida na cidade. Nota-se que seja também nos discursos que mais beiraram o nacionalismo, como uniformes nacionais e bandeiras, estão na verdade ligados ao movimento estético do *brazilcore*; Mesmo as representações vexolológicas, caberia abrir um aprofundamento da representação nacional. Conforme Matos e Monteiro, (1969) a história de nossa bandeira detém o mito político¹² de que suas cores representariam as simbologias das matas, rios e sol, ouro e o céu estrelado (ORTIZ, 1985), quando na realidade são as cores heráldicas das famílias imperiais Orleans e Bragança. Muitos heraldistas defendem que nossa bandeira deveria ser vermelha, uma vez que, linguisticamente, Brasil quer dizer “braseiro” e “cor de brasa”, devido as florestas de pau-brasil que havia em nosso território (MATOS; MONTEIRO, 1969). Para além do mito de representação da natureza nas cores de nossa bandeira, o corpus mostra que a natureza se encontra, nesse modelo de representação, reduzida a um cenário ruralista e turístico nas imagens do corpus. São vestígios ou novas facetas fora do imaginário de brasilidade eternamente vigente? Que imaginário seria esse? Kelly Kajihara (2008) e Bignami (2005), apontaram que, ao estudar a manutenção de imagem de Brasil pela Embratur, em termos de atratividade turística, a imagem do país de um modo geral se qualifica pelas seguintes categorias:

- Brasil Paraíso: idéia relacionada ao Éden, aos atrativos naturais e paisagísticos e as características descritas na carta de Pero Vaz de Caminha na época do descobrimento.
- Lugar de Sexo Frágil: relaciona-se a idéia de sensualidade, libertinagem e a beleza da mulher brasileira.
- Brasil do Brasileiro: incluem-se todas as características relacionadas ao povo brasileiro, como a musicalidade, a hospitalidade, a malandragem, a alegria, a cordialidade e a falta de preconceito.

¹² Hall (2006) defende que qualquer identidade nacional é falsa. Citando Schwartz, o autor salienta que “As identidades nacionais não são coisas com as quais nós nascemos, mas são formadas e transformadas no interior da representação. (...) Uma nação é uma comunidade simbólica e é isso que explica seu “poder para gerar um sentimento de identidade e lealdade” (HALL, 2006, pg 13).

- País do Carnaval: é a síntese do imaginário que associa o Brasil com grandes eventos na mídia, com o carnaval, o futebol e a música.
- Lugar do Exótico e do Místico: relaciona-se às manifestações religiosas, a cultura negra e indígena, aos ritos e rituais em geral. Tais estereótipos caracterizam o país de uma forma simplificada, na qual o diferencial do país fundamenta-se quase que exclusivamente na exuberância dos seus recursos naturais, na luxúria das festas e na beleza das mulheres.

Abre-se, assim, uma cisão entre a noção de brasilidade tradicional. Bem como uma cisão no discurso no corpus Flickr: a natureza real – flagrada na sua imensidão por pioneiros como Debret, Rugendas e Post – e a figurada nas imagens de Brasil contemporaneamente – nestas últimas, prevalecem as cenas litorâneas, de cânions ou de pastagens e plantações. As nove dimensões extraíram o discurso latente nas imagens, algo que muitos brasilianistas internacionais buscaram, de forma teórica.

5. CONCLUSÃO

Este estudo utilizou um corpus de imagens relacionadas ao Brasil para analisar as representações visuais do país, a fim de responder nossa questão inicial: *Quais discursos emanam da representação visual de Brasil?* O corpus foi compilado de forma criteriosa, levando em consideração a variedade e os objetivos da pesquisa. As imagens foram coletadas do site-portfolio Flickr por meio de um script desenvolvido pelo professor orientador. Em seguida, as características visuais de cada imagem foram etiquetadas usando o Google Cloud Visual IA por meio de scripts desenvolvido pelo professor orientador, e os rótulos foram processados estatisticamente, na busca por responder à pergunta empírica. A análise multidimensional, utilizando a técnica estatística da Análise Fatorial, foi realizada para identificar os fatores presentes nos dados e as dimensões subjacentes à variação do discurso subjacente presente nas imagens. Os escores de fator foram calculados para cada imagem e, em seguida, os escores médios foram calculados para cada nação. A análise estatística, incluindo o teste F de ANOVA, foi aplicada para verificar se havia diferenças significativas entre os escores médios das nações. A metodologia da Análise Multidimensional Lexical foi utilizada para identificar diferentes discursos relacionados à "brasilidade" nas imagens.

Cabe ainda refletirmos sobre em qual medida o presente estudo colabora com os estudos da Linguística de Corpus, estudos da cultura e as Humanidades Digitais. Como foi apresentado, lançou-se mão da fundamentação teórica da LC (BERBER SARDINHA, 2004, 2019), em justaposição aos estudos culturais digitais e de representação (HALL, 1997, 2014; MANOVICH, 2011, 2016).

Esta reflexão empírico-acadêmica examina criticamente os fundamentos teóricos e metodológicos presentes no conceito de *Cultural Analytics*, de Lev Manovich (2011, 2016), onde o autor compila imagens computacionais e as lê como manifestação cultural. Numa revisão de seu marco teórico, percebeu-se que os principais aspectos de cada abordagem (multimodalidade, cultura, história da arte, semiótica e análise cultural) são empreendimentos investigativos cooperativos entre si; e é na Linguística de Corpus, com sua característica empírica, que novas observações e considerações epistemológicas e metodológicas de pesquisa, que implicações para o estudo da cultura e da mensagem, ou discurso, são suscitados.

Baseando-se também numa revisão das mencionadas abordagens e suas produções bibliográficas, esta pesquisa visou fornecer uma análise estrita das diferentes perspectivas sobre a cultura, com foco na ideia de brasilidade, num estudo empírico, buscando contribuições para o campo mais amplo da análise cultural e multimodalidade, em linguística de corpus.

A Analítica Cultural como um avanço em Humanidades Digitais

A Analítica Cultural, desenvolvida por Lev Manovich, aprofunda os estudos em teoria da mídia e humanidades digitais, enfatizando o potencial de métodos computacionais para análise cultural em larga escala. Tal abordagem lê o cenário digital por meio de estudos baseadas em dados, buscando padrões e tendências em artefatos culturais (MANOVICH, 2011). Como agenda de pesquisa em LC e em Humanidades Digitais, o emprego de algoritmos e técnicas de visualização por computador, em ambas as ciências, poderá favorecer uma aproximação dos temas e metodologias das duas áreas, quantificando e potencializando conjuntos de dados culturais, revelando estruturas e correlações latentes (MANOVICH, 2015). O conceito de análise do nosso tempo, chamado de paisagem cultural contemporânea, permite uma análise macroscópica dos fenômenos culturais multimodais.

Os Estudos Decoloniais

Como poderíamos ler o Brasil apenas com as lentes de especialistas distantes da realidade histórica, contextual, local e cultural brasileira? Enraizados na teoria pós-colonial, os estudos decoloniais desafiam as epistemologias hegemônicas no Norte e procuram provocar uma reflexão das amarras, construções e manutenção dos legados coloniais (MIGNOLO, 2000). Essa ótica permeia a linguagem, preocupação da Linguística, ao se tencionar à reflexão o modo e do quanto os universalismos científicos, sua “imparcialidade” positivista, foram impostos aos povos do sul global. Ao se validar apenas saberes ocidentais, homogêneos e, como escreve Grada Kilomba (2019), “brancos”, se perpetua na realidade o silenciamento linguístico às multiplicidades não-europeias e não-

ocidentais. Assim, tais estudos estimulam o reconhecimento de diversos saberes e a prática de não termos bibliografias tão unicamente compostas por vozes do norte global, como forma de abertura das estruturas acadêmicas às epistemologias do sul (SMITH, 2012). Estudiosos decoloniais crescem para reafirmar a importância do conhecimento historicamente apagado de nossas bibliografias e enfatizam a necessidade de se envolver com contextos locais para entender a cultura de forma mais abrangente (GROSGOUEL, 2013; FANON, 2002). A abordagem dos estudos decoloniais destaca a heterogeneidade e a complexidade da cultura, reconhecendo sua interconexão com contextos históricos e sociais (QUIJANO, 2009)

A noção de cultura e de discurso de Stuart Hall

Os estudos culturais e de representação e discurso de Stuart Hall foram instrumentais na análise da produção e interpretação de mensagens culturais contidas neste trabalho. Seu modelo de codificação/decodificação ilumina o papel do poder, da ideologia e do discurso na formação dos significados histórico-culturais (HALL, 1980). A abordagem qualitativa e fundamentada etnograficamente de Hall examina a cultura como um local de luta, onde múltiplas interpretações coexistem e interagem (HALL, 1997). Essa perspectiva cultural faz uma análise com percepções diferenciadas, colocando em primeiro plano a interação dinâmica das representações culturais e os discursos (HALL, 1973).

Dessa maneira, destacando a jornada e arcabouços que serviram à essa pesquisa em Linguística de Corpus, a *Cultural Analytics* de Lev Manovich, os estudos decoloniais e as contribuições de Stuart Hall ofereceram perspectivas distintas e complementares sobre análise cultural. A abordagem computacional de Manovich é provocativa enquanto análise quantitativa em grande escala das tendências culturais, fornecendo informações valiosas sobre o cenário digital (MANOVICH, 2011). No entanto, graças aos avanços da escola de Linguística de Corpus de São Paulo, o GELC, as análises multimodais ganharam o frescor da combinação discursiva e extra disciplinar. Esta revisão e atitude possibilitaram, com a análise empírica das dimensões do corpus Flickr, em seus nove fatores, as

contribuições únicas expostas nesse trabalho e as interpretações multimodais extrapolando as fronteiras das análises discursivas e visuais. Baseada na metodologia desenvolvida pelo professor orientador Dr. Tony Berber Sardinha juntamente com o GELC, espera-se ter sido exposto aqui provocações e contribuições para o campo da análise cultural, multimodal e da linguística aplicada.

Em conclusão da parte teórica desta pesquisa, foi satisfatória mergulhar nas linhas e debates atuais sobre discurso, cultura, epistemologias e multimodalidade, nos estudos contemporâneos onde demonstraram que a linguagem é inerentemente multimodal, abrangendo não apenas palavras faladas e escritas, mas também pistas visuais, gestos e imagens (KRESS & VAN LEEUWEN, 2001). Esse reconhecimento provocou uma mudança de paradigma na linguística, visto que a integração da análise visual com o estudo linguístico permite uma compreensão mais abrangente dos padrões de comunicação, bem como dos aspectos culturais, sociais e cognitivos do uso da linguagem (JEWITT, 2017). Além disso, o surgimento da mídia digital e das tecnologias de comunicação intensificou a importância dos elementos visuais no discurso cotidiano (JEWITT, 2014). Na sociedade contemporânea, emojis, memes e infográficos tornaram-se componentes integrais da comunicação, transcendendo as barreiras linguísticas e facilitando as interações globais. Conseqüentemente, os pesquisadores em linguística devem abraçar os estudos de multimodalidade a pesquisa sobre visual comunicação, incorporando princípios de semiótica e análise de discurso multimodal para estudar o papel das imagens e outros elementos visuais em vários contextos de mídia (BATEMAN, 2016).

Já como considerações finais da análise empírica, esta pesquisa objetivou uma compreensão mais aprofundada das representações visuais do Brasil, destacando a importância do design de corpus e da análise multidimensional e visual para os estudos em LC. Em conclusão, este estudo busca contribuir para a compreensão das representações visuais do Brasil por meio da análise de um corpus de imagens. Os resultados mostraram diferenças significativas nas representações brasileiras, e a metodologia da Análise Multidimensional Lexical revelou diferentes discursos sobre a "brasilidade" nas imagens analisadas. E o trabalho reforçou a premissa de Hall (2006) de que toda e qualquer identidade nacional é falsa, construída. Tão falsa quanto as cores de nossa bandeira. E a

pesquisa proporcionou possibilidades empíricas de se analisar a construção desses discursos, no campo da multimodalidade.

No entanto, é importante considerar as limitações do estudo, como uma amostra específica, limitada no tempo, de imagens coletadas e a dependência de ferramentas de reconhecimento visual e análise estatística de acordo com as possibilidades do tempo presente. Ainda assim, que sirva de agenda para demais estudos e possibilidades em multimodalidade na Linguística de Corpus.

REFERÊNCIAS

ARTISTAS Viajantes. In: **ENCICLOPÉDIA Itaú Cultural de Arte e Cultura Brasileira**. São Paulo: Itaú Cultural, 2023. Disponível em: <http://enciclopedia.itaucultural.org.br/termo3778/artistas-viajantes>. Acesso em: 11 de maio de 2023. Verbete da Enciclopédia. ISBN: 978-85-7979-060-7.

BATEMAN, J. A.; SCHMIDT, K. H. **Multimodal film analysis: How films mean**. Routledge, 2008.

BATEMAN, J. A. Multimodal genre and evolving media. In J. Bateman & P. Wildfeuer (Eds.), **Text and image: A critical introduction to the visual/verbal divide** (pp. 181-200). Routledge, 2014.

BACKHAUS, Peter. **Linguistic landscapes: A comparative study of urban multilingualism in Tokyo**. Clevedon, UK: Multilingual Matters, 2007.

BAKHTIN, Mikhail. **A cultura popular na idade média e no renascimento: o contexto de François Rabelais**. São Paulo: Hucitec 2010.

BARTHES, Roland. **The Rhetoric of the Image: Image, Music, Text**. London: Fontana Press, 1977 [1964].

BEDNAREK, Monika. Analyzing pop culture. In: **The language of pop culture**. New York: Routledge, 2018. p. 255-264.

BERBER SARDINHA, Tony. O que é um corpus representativo? **DIRECT Papers**, n. 44, p. 1–26, 2000.

BERBER SARDINHA, Tony. **Linguística de Corpus**. Ed. Manole. São Paulo, 2004.

BERBER SARDINHA, Tony. 25 years later: Comparing Internet and pre-Internet registers. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia (Eds.). **Multi-Dimensional Analysis, 25 years on**. A tribute to Douglas Biber. Amsterdam/Philadelphia: John Benjamins, 2014. p. 81–105.

BERBER SARDINHA, Tony. **A corpus-based history of Applied Linguistic**. In: WORLD CONGRESS OF APPLIED LINGUISTICS (AILA). 2017.

BERBER SARDINHA, Tony. Dimensions of variation across Internet registers. **International Journal of Corpus Linguistics**, v. 23, n. 2, p. 125–157, 2018.

BERBER SARDINHA, Tony. Using Multi-Dimensional Analysis to Detect Representations of National Cultures. In: BERBER SARDINHA, Tony; VEIRANO PINTO, Marcia (Eds.). **Multi-Dimensional Analysis: Research Methods and Current Issues**. London; New York: Bloomsbury, 2019. p. 231–258.

BERBER SARDINHA, Tony. A historical characterisation of American and Brazilian cultures based on lexical representations. **Corpora**, v. 15, n. 2, p. 183–212, 2020.

BERBER SARDINHA, Tony. Discourse of academia from a multidimensional perspective. In: FRIGINAL, E.; HARDY, J. A. (Eds.). **The Routledge Handbook of Corpus Approaches to Discourse Analysis**. London; New York: Routledge, 2021. p. 298–318.

BERBER SARDINHA. A text typology of social media. **Register Studies**, v. 4, n. 2, p. 138–170, 27 out. 2022.

BERBER SARDINHA, Tony; VEIRANO PINTO, M. (Eds.). **Multi-Dimensional Analysis, 25 years on. A tribute to Douglas Biber**. Amsterdam/Philadelphia: John Benjamins, 2014.

_____. (Eds.). **Multi-Dimensional Analysis: Research methods and current issues**. New York: Bloomsbury Publishing, 2019.

BEZEMER, J.; KRESS, G. **Multimodality, learning and communication: a social frame**. New York: Routledge, 2016.

BIBER, D. **Variation across speech and language**. Cambridge: Cambridge University Press, 1988.

_____. **University Language**. Amsterdam/Philadelphia: John Benjamins, 2006.

BIBER, D.; CONRAD, S. **Register, Genre, and Style**. Cambridge: Cambridge University Press, 2009.

BIGNAMI, Rosana. **A imagem do Brasil no turismo**. 2ª Edição. São Paulo: Ed. Aleph, 2005.

BRAIT, Elisabeth. Bakhtin: **Conceitos-chave**. São Paulo: Contexto, 2005.

BROGINI, Arianne A. B. **Representações contemporâneas da sustentabilidade: uma análise multidimensional lexical discursiva como contribuição para o Portal Multimodal/Multilíngue para Avanço da Ciência Aberta nas Humanidades**. [s.l.] Pontifícia Universidade Católica de São Paulo, 28 mar. 2023.

CAMINHA, Pero Vaz. Carta de Pêro Vaz de Caminha. 1 de Maio de 1500. Portugal, Torre do Tombo, Gavetas, Gav. 15, mç. 8, n.º 2. Disponível em WWW: <URL: http://purl.pt/162/1/brasil/obras/carta_pvcaminha/index.html> Acesso em: Setembro de 2021.

CARNEIRO, Sueli, 2006. **A Construção do Outro como Não-Ser como fundamento do Ser**. Tese (Doutorado em Educação). São Paulo, SP: Universidade de São Paulo, FEUSP. [Acesso em 27 junho 2022]. Disponível em: <https://repositorio.usp.br/item/001465832>.

CLARKE, Isobelle; GRIEVE, Jack. **Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018**. *PloS one*, v. 14, n. 9, p. e0222062, 25 set. 2019.

COSTA, Gil. Anos 1980, 'visualidade amazônica': o desejo pela Amazônia na arte brasileira. In: **Anais do XXXVIII Congresso do CBHA**, 2018. pp. 530-543. Disponível em: <http://www.cbha.art.br/coloquios/2018/anais/pdfs/04%20Gil%20Vieira%20Costa.pdf>

DELFINO, Maria Cláudia. **More than words: análise multimodal multidimensional da música popular em língua inglesa**. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, 2022.

FANON, Frantz. **Os condenados da Terra**. Minas Gerais: Editora UFJF, 2002.

FILLMORE, Charles J.; BAKER, Collin F. **Frame semantics for text understanding**. Proceedings of WordNet and Other Lexical Resources Workshop, NAACL. 2001.

GABRIELLI, Lourdes; CAMELO, Francisco. O potencial visual dos sinais na comunicação: a poesia visual e a escrita logográfica marajoara: Kusiwa, do abstrato ao icônico. In: **Revista Sentidos da Cultura** Revista Semióticas, UFPA. V. 08 N. 15, 2021. pp. 53-65.

GADELHA, Hayle. The art of diplomacy. In: **The art of Diplomacy: Brazilian Modernism Painted for war**. Londres: Embaixada do Brasil no Reino Unido, 2018. pp. 43-55.

GATTI, Bernadete A. **Grupo focal na pesquisa em Ciências Sociais e Humanas**. Brasília: Liber Livro, 2005.

GOMBRICH, Ernst H. **A história da arte**. Rio de Janeiro: LTC, 2013.

GREENBAUM, Thomas L. **The Handbook for Focus Group Research**. Thousand Oaks: SAGE, 1998.

HAJER, Marteen A. **The Politics of Environmental Discourse: Ecological Modernization and the Policy Process**. New York, NY: Oxford University Press, 1997.

GROSFUGUEL, R. The structure of knowledge in Westernized universities: Epistemic racism/sexism and the four genocides/epistemicides of the long 16th century. **Human Architecture: Journal of the Sociology of Self-Knowledge**, 11(1), 2013. Pp. 73-90.

HALL, Stuart. **Encoding and Decoding in the Television Discourse**. Birmingham: Centre for Contemporary Cultural Studies, 1973.

_____. **Representation: Cultural Representations and Signifying Practices**. Sage Publications Ltd., 1997.

_____. **The Question of Cultural Identity**. In: S. HALL; D. HELD; T. MCGREW (Eds.). **Modernity and Its Futures**. Milton Keynes. Cambridge: Open University Press, 1992.

_____. **Cultura e representação**. Rio de Janeiro: Apicuri; PUC-Rio, 2016.

HALLIDAY, M. **Language as Social Semiotic: The Social Interpretation of Language and Meaning**. London: Edward Arnold, 1978.

JEWITT, C. An introduction to multimodality. In: **The Routledge Handbook of Multimodal Analysis**. Routledge, 2014. pp. 17-30.

JEWITT, C. Multimodal methods for researching digital technologies. In: **The Routledge Handbook of Language and Digital Communication**. Routledge, 2017. pp. 49-64.

KAJIHARA, Kelly Akemi; **A imagem do Brasil no exterior: análise do material de divulgação oficial da EMBRATUR, desde 1966 até os dias atuais**. USP, São Paulo. 2008

KAUFFMANN, Carlos H. **Linguística de corpus e estilo: análises multidimensional e canônica na ficção de Machado de Assis**. 277 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020. Disponível em: <<https://tede2.pucsp.br/handle/handle/23128>>.

KELLY-HOLMES, Helen. 2005. **Publicidade como comunicação multilíngue**. Basingstoke: Palgrave Macmillan.

KILOMBA, Grada. **Memórias da Plantação**. Rio de Janeiro: Cobogó, 2019.

KRESS, G. **Multimodality: A social semiotic approach to contemporary communication**. Routledge, 2010

KRESS, G.; VAN LEEUWEN, T. **Multimodal discourse: The modes and media of contemporary communication**. Oxford University Press, 2001.

MACHIN, D.; MAYR, A. **How to do critical discourse analysis: A multimodal introduction**. Sage Publications Ltd, 2012.

MANOVICH, Lev. **The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics**. Journal of Cultural Analytics. Nova York: May, 2016. PP. 123-147.

_____. **Trending: The promises and the challenges of big social data**. Debates in the Digital Humanities, 2011. Pp. 460-475.

_____. Cultural analytics: Analysis and visualization of large cultural data sets. **International Journal of Cultural Analytics and Social Theory**, 1(1), 2015. pp. 3-20.

_____. **Trending: The promises and the challenges of big social data**. Debates in the Digital Humanities, 2011. Pp. 460-475.

_____. Cultural analytics: Analysis and visualization of large cultural data sets. **International Journal of Cultural Analytics and Social Theory**, 1(1), 2015. pp. 3-20.

MATOS, Gastão; MONTEIRO, Luís. **Heráldica**. Ed. Verbo. Campinas, 1969

MAYER, C. **O que e como escrevemos na web: um estudo multidimensional de variação de registro em língua inglesa**. 129f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020. Disponível em: <<https://tede2.pucsp.br/handle/handle/21448>>.

MOITA LOPES, Luis P. Uma Linguística aplicada mestiça e ideológica: interrogando o campo do linguista aplicado. In: MOITA LOPES, Luis P. (Org.). **Por uma Linguística Indisciplinar**. São Paulo: Parábola Editorial, 2006. p.13-43.

MORIN, Edgar. **Introdução ao pensamento complexo**. Porto Alegre: Sulina, 2005.

MIGNOLO, W. D. **Local Histories/Global Designs: Coloniality, Subaltern Knowledges, and Border Thinking**. Princeton University Press, 2000.

NYHAN, Julianne; PASSAROTTI, Marco (Eds.). **One Origin of Digital Humanities: Fr Roberto Busa in His Own Words**. Springer Nature Switzerland, 2019.

ORTIZ, Renato. **Cultura brasileira e identidade nacional**. São Paulo: Brasiliense, 1985.

PANOFSKY, Erwin. O significado nas artes visuais. São Paulo: Perspectiva, 2017.

PENNYCOOK, Alastair. Uma linguística Aplicada Transgressiva. In: MOITA LOPES, Luis Paulo. (Org.). **Por uma Linguística Indisciplinar**. São Paulo: Parábola Editorial, 2006. p.67-84.

PENNYCOOK, Alastair; OTSUJI, Emi. **Metrolinguismo: Linguagem na cidade**. New York: Routledge, 2015.

PINHEIRO, Osmar. A visualidade amazônica. In: FUNARTE. **As artes visuais na Amazônia: reflexões sobre uma visualidade regional**. Rio de Janeiro: Funarte; Belém: SEMEC, 1985. p. 89-100.

QUIJANO, Aníbal. Colonialidade do Poder e Classificação Social. Em: B. S. Santos & M. P. Meneses (orgs.), **Epistemologias do sul**. Coimbra: Almedina, I, 2009. pp.73-117.

RAMOSE, Mogobe. Globalização e Ubuntu. Em: B. S. Santos & M. P. Meneses (orgs.), **Epistemologias do sul**. Coimbra, Almedina p.135-170.

ROMEIRO, Yara M. T. D. **A linguagem verbal das artes visuais: uma análise multidimensional do discurso sobre a fotografia de Sally Mann**. 182 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) - Pontifícia Universidade Católica de São Paulo, São Paulo, 2020. Disponível em: <<https://sapientia.pucsp.br/handle/handle/23462>>.

SANTAELLA, Lúcia. **O que é semiótica**. São Paulo: Contexto, 2002

SHEPHERD, Tania G.; SALIÉS, Tania G. (Eds.). **Linguística da Internet**. São Paulo: Contexto, 2013.

SMITH, L. T. **Decolonizing methodologies: Research and Indigenous peoples**. Zed Books, 2012.

SOUZA, Rafael W. F. de. **As marcas lexicais da discriminação étnico-racial entre negros e brancos**: um estudo da linguística de corpus. Dissertação de mestrado, LAEL. São Paulo: Pontifícia Universidade Católica de São Paulo, 2020.

VAN LEEUWEN, T. **The language of colour: an introduction**. London: Routledge, 2011.

VEIGA, Alexandre T. **As dimensões da fé: sete religiões mundiais em uma análise multidimensional lexical**. Tese de doutoramento, LAEL. São Paulo: Pontifícia Universidade Católica de São Paulo, 2021.

WAY, Lyndon C. S. ; MCKERRELL, Simon. **Music as Multimodal Discourse: Semiotics, Power and Protest**. London: Bloomsbury, 2017.

ANEXOS

ANEXO I – CORPUS FLICKR

O autor disponibilizou para consulta a lista completa das imagens utilizada nesta pesquisa, bem como uma pasta contendo os exemplos mais significativos de cada fator, dentre as 5.833 imagens que compuseram o corpus Flickr.

Figura 65 - Corpus Flickr, com 5833 itens JPG.

```

5796 005889 britain 9588276922_d55bf1e6b3_b.jpg
5797 005890 britain 49629988656_87f64769ed_b.jpg
5798 005891 britain 5078262050_fa82bf47ba_o.jpg
5799 005892 britain 27090135004_fc3a945312_b.jpg
5800 005893 britain 6271885452_4a178c9b57_b.jpg
5801 005894 britain 6582573839_3cFb4be429_b.jpg
5802 005895 britain 4619257747_4a2e6cf809_b.jpg
5803 005896 britain 7018642547_3555cea357_b.jpg
5804 005897 britain 46971871482_181094d195_o.jpg
5805 005898 britain 7594477166_c4fa2cacd5_b.jpg
5806 005899 britain 8203533121_rec27139a2_b.jpg
5807 005900 britain 7613792529_d898f89e52_b.jpg
5808 005901 britain 8136289795_e8995d5e4a_b.jpg
5809 005902 britain 24287545271_fac190acc4_o.jpg
5810 005903 britain 8055026668_068d6c5aff_b.jpg
5811 005904 britain 4619872414_0096fd87ae_b.jpg
5812 005905 britain 4568842112_bec9e42bee_o.jpg
5813 005906 britain 8100767040_05f3bec595_b.jpg
5814 005907 britain 5926906221_c3242621f1_b.jpg
5815 005908 britain 7960552966_79b9a7a3df_b.jpg
5816 005909 britain 13433797345_77954d82f8_o.jpg
5817 005910 britain 5926906679_0b5d962ae1_b.jpg
5818 005911 britain 7573952629_7e54430af3_b.jpg
5819 005912 britain 8136289927_e81b33e883_b.jpg
5820 005913 britain 8204603984_e62bacl279_b.jpg
5821 005914 britain 8055026292_f4e4f8c0f0_b.jpg
5822 005915 britain 51694040320_870fa80981_o.jpg
5823 005916 britain 51215894613_be7a2b0c2c_b.jpg
5824 005917 britain 7546998498_fffc45f584_b.jpg
5825 005918 britain 8055046247_b9847100fe_b.jpg
5826 005919 britain 5916797014_e85e22d91a_o.jpg
5827 005920 britain 5154438482_fc9c8492fb_o.jpg
5828 005921 britain 7855293176_d4aecdd6b7_b.jpg
5829 005922 britain 67470389_3a947bb8d2_o.jpg
5830 005923 britain 38503204260_4ba784499b_o.jpg
5831 005924 britain 27886481111_416996cf02_b.jpg
5832 005925 britain 5214885466_37a7bf101c_b.jpg
5833 005926 britain 47301705081_47aee164af_b.jpg
5834

```

Fonte: Captura de tela pelo autor.

Confira a lista completa das imagens em (Abrir em Notepad++ ou bloco de notas similar):

<https://drive.google.com/file/d/1KwTGZFOoh2C1ttCDHmDCAPZbKZP798a/view?usp=sharing>

Banco de imagens em JPG:

https://drive.google.com/drive/folders/1URBikk2TqXdBIDzDxrc24IMcd1aX4-05?usp=drive_link

ANEXO II – ESCORE DO SAS, GLM:

Capturas de tela que asseguram a precisão da ANOVA, com valor de R quadrado em cada amostra:

GLM for dataset = luiz_br_vs_non-br_scores_country f1**The GLM Procedure**

Class Level Information		
Class	Levels	Values
country	4	br de gb us

Number of Observations Read	3870
Number of Observations Used	3870

GLM for dataset = luiz_br_vs_non-br_scores_country f1

The GLM Procedure

Dependent Variable: fac1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	219375.6865	73125.2288	1139.44	<.0001
Error	3866	248107.0629	64.1767		
Corrected Total	3869	467482.7494			

R-Square	Coeff Var	Root MSE	fac1 Mean
0.469270	339.5696	8.011035	2.359173

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	3	219375.6865	73125.2288	1139.44	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	3	219375.6865	73125.2288	1139.44	<.0001

GLM for dataset = luiz_br_vs_non-br_scores_country f1

The GLM Procedure

Level of country	N	fac1	
		Mean	Std Dev
br	997	-1.5125376	2.7548958
de	984	-2.1402439	2.9205565
gb	948	15.5696203	15.3851780
us	941	-2.1424017	2.9181964

GLM for dataset = luiz_br_vs_non-br_scores_country f2

The GLM Procedure

Class Level Information		
Class	Levels	Values
country	4	br de gb us

Number of Observations Read	3870
Number of Observations Used	3870

GLM for dataset = luiz_br_vs_non-br_scores_country f2

The GLM Procedure

Dependent Variable: fac2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	15029.3229	5009.7743	107.90	<.0001
Error	3866	179492.5748	46.4285		
Corrected Total	3869	194521.8977			

R-Square	Coeff Var	Root MSE	fac2 Mean
0.077263	191.0009	6.813846	3.567442

Source	DF	Type I SS	Mean Square	F Value	Pr > F
country	3	15029.32289	5009.77430	107.90	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
country	3	15029.32289	5009.77430	107.90	<.0001

GLM for dataset = luiz_br_vs_non-br_scores_country f2**The GLM Procedure**

Level of country	N	fac2	
		Mean	Std Dev
br	997	5.35205617	8.92682368
de	984	4.48780488	7.05059853
gb	948	0.20042194	3.22213828
us	941	4.10626993	6.63848972

ANEXO III – POSTAGENS NO BLOG

O aluno executou, ao longo de sua pesquisa, algumas publicações em política de Ciência Aberta, divulgando o andamento e a evolução temporal de sua pesquisa. Abaixo, uma curta amostra dessas publicações, disponíveis em: <https://cienciaaberta.org/author/luiz-carlos-zeferino/>



(<http://cienciaaberta.org>)

dezembro 1, 2022

Brasilidade Vs Brazilcore

Recentemente vi romper na aurora das redes sociais uma linguagem fashionista, batizada internacionalmente de “Brazil Core”, também chamado de Brazilcore e Brazil Aesthetics. Um exemplo abaixo:



Nem se sabe mais se é moda ou movimento político (<https://falauniversidades.com.br/brazilcore-moda-ou-movimento-politico/>). Imediatamente levantei uma enquete, ainda coletando respostas, pois me inquietou muitíssimo. Talvez eu aborde o tema num artigo, semestre que vem...

Confira aqui a enquete e participe respondendo sua opinião, se desejar: <https://forms.gle/9sdaewF46qPtpSBv9>

<https://cienciaaberta.org/2022/12/01/brasilidade-vs-brazilcore/>

01/08/2023, 23:27 Apresentação em Portugal e Processamento recebido, agora é dissertar!
– Portal da Ciência Aberta

março 6, 2023

Apresentação em Portugal e Processamento recebido, agora é dissertar!

Graças ao meu orientador, recebi o corpus do Flickr todo processado. Agora basta escrever. Foi triste abrir mão da minha ideia original de pesquisa, mas ela se tornará um artigo, aguardem!

Neste mês meu trabalho foi apresentado na Universidade do Minho, em Braga, Portugal. Uma experiência a ser lembrada, claro. Thanks, 1º DDHUM! Thanks, Tony Berber Sardinha!

 Luiz Zeferino (<https://cienciaaberta.org/author/luiz-carlos-zeferino/>)

 0

 Limitação de Código, Uso do ChatGPT e soluções improvisadas
(<https://cienciaaberta.org/2023/02/17/limitacao-de-codigo-uso-do-chatgpt-esolucoes-improvisadas/>)

Teorias utilizadas para a seleção de per s 

<https://cienciaaberta.org/2023/03/06/apresentacao-em-portugal-e-processamento-recebido-agora-e-dissertar/>