



## DIMENSÕES DE VARIAÇÃO LEXICAL E ACÚSTICA NA MÚSICA POPULAR EM INGLÊS: UM ESTUDO BASEADO EM CORPUS<sup>1</sup>

### LEXICAL AND ACOUSTIC DIMENSIONS OF VARIATION IN POPULAR MUSIC IN ENGLISH: A STUDY BASED IN CORPUS

Maria Claudia Nunes Delfino<sup>2</sup>

Tony Berber Sardinha<sup>3</sup>

Joseph Glynn Collentine<sup>4</sup>

**Resumo:** Embora a música popular tenha sido objeto de estudo linguístico, tem havido pouco interesse na descrição da canção do ponto de vista linguístico e musical simultaneamente. A maioria dos estudos descritivos foca no modo verbal da produção musical, isto é, o texto da composição (a letra da música). Neste estudo, buscamos realizar uma descrição tanto do texto escrito da canção popular em inglês quanto de sua manifestação musical, de tal modo a proporcionar uma visão holística desse produto cultural. Para tanto, empregamos uma abordagem baseada na Linguística de Corpus, por meio da qual foi possível coletar e analisar uma grande amostra de canções em inglês, incluindo mais de 200 mil letras de música e mais de 97 mil canções indexadas acusticamente. A metodologia baseou-se na Análise Multidimensional, uma abordagem baseada em corpus que permite identificar as dimensões de variação subjacentes a uma determinada variedade linguística. Em relação ao texto da composição musical, quatro dimensões foram identificadas, as quais refletem a predominância de determinados discursos. Em relação ao componente acústico, foram identificadas três dimensões, cada uma correspondendo a uma determinada musicalidade. A pesquisa ilustra a possibilidade de descrição linguística e acústica em larga escala por meio de metodologias baseadas em corpus. De modo geral, os resultados mostram que a música popular em inglês, embora extremamente variada, pode ser resumida em torno de quatro padrões textuais e três padrões musicais.

**Palavras chave:** Linguística de Corpus; Música popular; Análise Multidimensional.

**Abstract:** Although popular music has received attention in linguistics, there has been little interest in describing music from both a linguistic and a musical standpoint simultaneously. Most linguistic studies focus on the verbal mode of the musical production, that is, the text of the composition (the lyrics). In this study, we seek to conduct a description of both the written text and the acoustic indices of popular music, in such a way as to provide a holistic view of this cultural product. To do so, we employed a corpus approach, through which it was possible to collect and analyze a large sample of English songs, including more than 200,000 lyrics and more than 97,000 acoustically indexed songs. The methodology was based

---

<sup>1</sup> Os autores agradecem o apoio financeiro das seguintes organizações: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Processo 2022/05848-7; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Processos 310140/2021-8 e 420180/2022-2; Pontifícia Universidade Católica de São Paulo, PIPEq 21972; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Processos PROSUP/PROSUC 88887.177403/2018-00, PDSE 88881.362173/2019-01.

<sup>2</sup> Professora da Faculdade de Tecnologia de Praia Grande e aluna de Pós-Doc na Pontifícia Universidade Católica de São Paulo – PUC-SP. [maria.delfino@fatec.sp.gov.br](mailto:maria.delfino@fatec.sp.gov.br)  
Orcid: <https://orcid.org/0000-0003-1976-8725>

<sup>3</sup> Professor da Pontifícia Universidade Católica de São Paulo – PUC-SP. [tonycorpuslg@gmail.com](mailto:tonycorpuslg@gmail.com)  
Orcid: <https://orcid.org/0000-0001-8815-1521>

<sup>4</sup> Professor da Northern Arizona University – NAU. [Joseph.Collentine@nau.edu](mailto:Joseph.Collentine@nau.edu)  
Orcid: <https://orcid.org/0000-0002-1815-3449>

on Multidimensional Analysis, a corpus-based approach that enables the identification of the dimensions of variation underlying a given linguistic variety. Regarding the lyrics, four dimensions were identified, each reflecting the predominance of particular discourses. Regarding the acoustic component, three dimensions were identified, each corresponding to a particular musical pattern. This paper illustrates the possibility of large-scale linguistic and acoustic description using corpus-based methodologies. Overall, the results show that popular music in English, although extremely varied, can be summarized around four textual patterns and three musical patterns.

**Key words:** Corpus Linguistics; Popular music; Multi-Dimensional Analysis

## 1. INTRODUÇÃO

A música popular em língua inglesa tem sido objeto de diversos estudos nas mais variadas áreas, como Musicologia (MIDDLETON, 1993; MOORE, 2021), Estudos Culturais (WALL, 2013), Sociologia (BENNETT, 2008), Semiótica (DUNBAR-HALL, 1991) e Linguística (BARBOSA, 2012; BÉRTOLI, 2002, 2010, 2014; DELFINO, 2016; WERNER, 2012; 2021), entre outras. Além de ser um produto comercial rentável, também é um produto cultural que contribui para moldar o comportamento e a vida em sociedade (VLADI, 2010).

Há pelo menos três tendências gerais na literatura. A primeira é a existência de um número reduzido de pesquisas empíricas sobre o discurso da música, visto que existe uma predominância de trabalhos de natureza teórica (BARBOSA, 2012; BENNETT, 2008). A segunda é o fato de haver certo reducionismo, na medida em que existe uma desvinculação do componente verbal (a letra das músicas) dos seus elementos fundamentais (o ritmo, a melodia e a harmonia), complementado pelo fato de a letra da música receber mais atenção do que a melodia (WALL, 2013, p. 128). Por fim, a terceira tendência diz respeito ao emprego, entre os estudos empíricos, de amostras reduzidas de dados, formadas por poucos exemplares de canções, o que reduz o alcance do estudo e, por conseguinte, a possibilidade de generalizar os resultados (DELFINO, 2016; WERNER, 2012, 2021).

Tendo em vista essas lacunas, a presente pesquisa tem como objetivo o desenvolvimento de uma análise (1) multimodal, isto é, que envolva mais de um modo de expressão, neste caso tanto o componente verbal (a letra da música) quanto o componente acústico (a sonoridade da voz e dos instrumentos musicais); (2) multidimensional, isto é, que possibilite a identificação de várias dimensões capazes de revelar o discurso presente no registro música em língua inglesa, tanto dimensões acústicas como dimensões lexicais, ambas capazes de revelar padrões coocorrentes de características; e (3) empírica com emprego de *big data*, isto é, baseada em uma grande amostra representativa da população de canções populares em inglês.

A fim de atingir os objetivos elencados, recorreremos à Linguística de Corpus, uma área dos Estudos Linguísticos que se dedica à coleta e exploração de grandes amostras de textos falados, escritos, acústicos ou visuais, registrados na forma de arquivos de computador (BERBER SARDINHA, 2000, 2004; BIBER; CONRAD; REPPEN, 1998; MCENERY; WILSON, 1996; BERBER SARDINHA, 2004). Tais amostras, chamadas de corpora, passam por análises que buscam identificar os padrões de uso da linguagem. Padrões são entendidos como conjuntos de características estatisticamente recorrentes, que revelam sentidos associados ao uso (BERBER SARDINHA, 2004; HOEY, 2005; HUNSTON; FRANCIS, 2000; SINCLAIR, 1966; 1988; 1991; SINCLAIR; JONES; DALEY, 1970/2004).

No âmbito da Linguística de Corpus, a Análise Multidimensional apresenta-se como uma vertente metodológica que emprega procedimentos estatísticos multivariados a fim de identificar conjuntos de características linguísticas correlacionados que são

compartilhados entre textos do corpus. Esses conjuntos são identificados por meio de análise fatorial, cujos fatores resultantes, ao serem interpretados comunicativamente, assumem o estatuto de dimensão (BERBER SARDINHA, 2000, BERBER SARDINHA e VEIRANO, PINTO, 2014, 2019; BIBER, 1986, 1988, 1995). O foco da Análise Multidimensional é descrever a variação entre variedades textuais situacionalmente definidas, chamadas de ‘registros’ (BIBER, 1994, 1995, 2012, 2019, BIBER; CONRAD, 2009), como por exemplo conversação (BIBER, 2004, 2008), artigos acadêmicos (GRAY, 2015; HARDY, 2015), mensagens de redes sociais (BERBER SARDINHA, 2014, 2018, 2022) etc. Na medida em que a variação entre os registros é descrita por meio de dimensões, podemos considerar a dimensão como o parâmetro responsável pela variação entre os textos.

A primeira Análise Multidimensional, realizada por Douglas Biber nos anos 1980 (BIBER, 1988), enfocou a variação entre registros da língua inglesa, com base em características gramaticais. O estudo identificou cinco dimensões principais: (1) Produção envolvida versus informacional, (2) Propósitos narrativos versus não-narrativos, (3) Referência explícita versus dependente de situação, (4) Persuasão explícita e (5) Informação abstrata versus não-abstrata. Por exemplo, a primeira dimensão compreende dois polos, um referente à produção envolvida e outro à produção informacional. Cada polo corresponde a um conjunto distinto de características linguísticas, distribuídas de tal forma que os textos que possuem as características de um polo provavelmente terão pouca incidência de características do outro polo. A incidência de cada dimensão em cada texto é calculada, resultando em um escore de texto por dimensão, tornando assim possível mensurar os textos e os registros do corpus em cada dimensão. Como os diferentes registros da língua inglesa são ranqueados ao longo de cada dimensão, pode-se verificar quais os registros mais tipicamente associados a cada um dos polos. Por exemplo, a conversação face-a-face comporta características típicas do polo positivo, que são as formas pronominais de primeira e segunda pessoa, verbos no tempo presente, formas contraídas, entre outras. A escrita acadêmica, por sua vez, emprega características linguísticas do polo informativo, tais como formas nominais, preposições e adjetivos em posição atributiva.

As dimensões de variação da língua inglesa são regularmente empregadas como base para a descrição de registros ainda não analisados multidimensionalmente. Esse tipo de Análise Multidimensional, em que registros são comparados a dimensões preexistentes, é chamada de análise aditiva (BERBER SARDINHA, 2019). Na literatura de Análise Multidimensional, existem pesquisas sobre música pop, como Werner (2021) e Delfino (2016), que realizaram análise aditiva desse registro com base nas dimensões de variação de registro do inglês de Biber (1988). BÉRTOLI (2014), diferentemente dos estudos anteriores, extraiu novas dimensões de variação com base em um grande corpus de música pop em inglês compreendendo o período entre 1940 e 2009. Embora esses estudos tenham tido objetivos diferentes ao realizar Análise Multidimensional, eles têm em comum o fato de buscarem uma caracterização funcional da música popular em inglês, a partir da descrição gramatical das letras de música.

Na presente pesquisa, no entanto, empregamos uma vertente diferente da Análise Multidimensional, que se destina à identificação de grupos correlacionados de itens lexicais, compartilhados entre os textos (BERBER SARDINHA, 2023). Essa vertente, conhecida como Análise Multidimensional Lexical, permite identificar textos que compartilham conjuntos lexicais correlacionados. Tais conjuntos lexicais, os quais são identificados por meio de análise fatorial, assim como na Análise Multidimensional de base gramatical, são interpretados como índices que ajudam o analista a chegar aos temas

ou discursos subjacentes (BERBER SARDINHA, 2020, 2021a; CLARKE; MCENERY; BROOKES, 2021; FITZSIMMONS-DOOLAN, 2014; 2019).

Vários estudos empregaram a Análise Multidimensional para identificar discursos em domínios diversos, como por exemplo, o discurso religioso (VEIGA, 2021), mídias sociais (MILANEZ; KAUFFMANN; DELFINO; BERTOLI *et al.*, 2021), representações nacionais (BERBER SARDINHA, 2020), artes visuais (ROMEIRO, 2020), o estilo de Machado de Assis (BERBER SARDINHA; KAUFFMANN, 2020), os discursos da Linguística Aplicada no Brasil (DELFINO; ARAÚJO; BERBER SARDINHA, 2018) e no exterior (BERBER SARDINHA, 2021a, 2023), bem como o inglês acadêmico (ZUPPARDI; BERBER SARDINHA, 2020), entre outros. Por exemplo, BERBER SARDINHA (2020) identificou dimensões lexicais na coletânea eletrônica Google Books, a fim de detectar representações de brasilidade e americanidade de 1800 a 2008. Para tanto, focou nas palavras que ocorriam em seguida dos adjetivos *Brazilian* ou *American* nos textos. Para ilustrar, em relação a ‘Brazilian’, foram detectadas cinco dimensões, cada uma correspondendo a uma representação distinta da identidade brasileira. Por exemplo, a primeira dimensão refletia a representação de Brasil do ponto de vista de sua economia e política, o que era marcado lexicalmente pelo uso constante de vocábulos como *economic*, *development*, *industrial*, *military* e *political*. Essa dimensão tornou-se dominante a partir dos anos 1960 até os anos 1990, atingindo seu ápice nos anos 1970. Essa visão de um país em industrialização com uma cena política turbulenta contrasta com a representação dominante do país no século XIX até início do século XX, a qual corresponde à dimensão 5, formada por itens lexicais como *forest*, *river*, *mine*, *province* e *steamer*, que dá vazão a uma representação do Brasil como uma nação selvagem e provinciana.

No âmbito da Análise Multidimensional lexical, entendemos discurso como *uma atividade social que produz o sentido, situada historicamente e que pode ser traçada a partir de escolhas lexicais*<sup>5</sup> (BERBER SARDINHA, 2021a, p. 299). Tal entendimento se apoia em concepções de discurso como a de Baker e McEnery (2015, p. 5), segundo a qual os discursos são *maneiras de olhar o mundo, construir objetos e conceitos de determinadas maneiras, representando a realidade*<sup>6</sup>, na de Burr (1995, p. 48), segundo a qual discursos são *o conjunto de significados [...], representações [...], afirmações que ao estarem juntos produzem uma visão específica de eventos*<sup>7</sup>, na de HAJER (1993, p. 44), que entende discursos como *um conjunto de ideias, conceitos e categorizações que são produzidas, reproduzidas e transformadas em um determinado grupo de práticas pelas quais o significado é dado para realidades físicas e sociais*<sup>8</sup>, bem como na de HALL (1992, p. 201), para quem discursos são:

*um grupo de enunciados que fornecem uma base linguística para falar – ou seja, uma forma de representar – um determinado tipo de conhecimento acerca de um tópico. Quando os enunciados*

---

<sup>5</sup> No original: A historically situated, meaning-making social activity that can be traced in part through lexical choices.

<sup>6</sup> No original: Ways of looking at the world, of constructing objects and concepts in certain ways, of representing reality

<sup>7</sup> No original: the set of meanings, [...] representations, [...] statements and so on that in some way together produce a particular version of events.

<sup>8</sup> No original: an ensemble of ideas, concepts, and categorizations that are produced, reproduced and transformed in a particular set of practices and through which meaning is given to physical and social realities

se realizam no âmbito de um determinado discurso, este torna possível a construção do tema de uma determinada forma. Também limita as outras formas em que o tópico pode ser construído<sup>9</sup>.

## 2. METODOLOGIA

A fim de possibilitar a descrição dos modos verbal e acústico das canções, foram empregados dois conjuntos de dados. O primeiro consiste em um corpus textual de letras de música em inglês, chamado CoLiE lyrics, coletado especialmente para esta pesquisa a partir do site *lyrics.com* em junho de 2019. A seguir, esses arquivos foram curados manualmente. Dois critérios principais para a curadoria do conteúdo nortearam essa etapa: (a) manter somente de letras de música em língua inglesa; e (b) usar apenas a primeira gravação nas músicas com mais de uma gravação. Dos 236.756 arquivos originários da coleta inicial, a curadoria eliminou mais de 30 mil arquivos. O corpus resultou em 201.235 canções, compreendendo 103.809.650 ocorrências (tokens) e 164.367 vocábulos (types).

O corpus textual foi etiquetado em categorias semânticas usando o etiquetador USAS (sigla de UCREL *Semantic Analysis System*), desenvolvido na Universidade de Lancaster (<http://ucrel-api.lancaster.ac.uk/usas/tagger.html>). Ele se baseia na classificação do léxico em categorias distintas, de tal modo que as palavras classificadas em uma mesma categoria compartilhem um campo semântico semelhante. Por exemplo, a categoria *education* abarca termos do âmbito da educação, como *class*, *school*, *course*, *board*, *college*, *university*, *private*, *classroom*, entre outras.

A taxonomia do USAS compreende três níveis, totalizando 453 categorias, distribuídas em níveis hierárquicos. O nível superior contém 21 categorias, conforme mostrado no Quadro 1. Cada uma das categorias superiores é subdividida em subcategorias; por exemplo, a categoria ‘General and abstract terms’ possui 14 subcategorias, conforme mostra o Quadro 2. Essas subcategorias, por sua vez, possuem uma divisão interna. Por exemplo, a subcategoria A5, *Evaluation*, compreende 12 subcategorias próprias (vide Quadro 3). Nesta pesquisa, empregamos os dois primeiros níveis da taxonomia.

Quadro 1. Categorias superiores do etiquetador semântico USAS

<b>Categoria</b>	<b>Descrição</b>
A	GENERAL & ABSTRACT TERMS
B	THE BODY & THE INDIVIDUAL
C	ARTS & CRAFTS
E	EMOTIONAL ACTIONS, STATES & PROCESSES
F	FOOD & FARMING
G	GOVT. & THE PUBLIC DOMAIN
H	ARCHITECTURE, BUILDINGS, HOUSES & THE HOME
I	MONEY AND COMMERCE
K	ENTERTAINMENT, SPORTS & GAMES

<sup>9</sup> No original: a group of statements which provide a language for talking about — i.e. a way of representing — a particular kind of knowledge about a topic. When statements about a topic are made within a particular discourse, the discourse makes it possible to construct the topic in a certain way. It also limits the other ways in which the topic can be constructed

L	LIFE AND LIVING THINGS
M	MOVEMENT, LOCATION, TRAVEL & TRANSPORT
N	NUMBERS & MEASUREMENT
O	SUBSTANCES, MATERIALS, OBJECTS & EQUIPMENT
P	EDUCATION
Q	LINGUISTIC ACTIONS, STATES & PROCESSES
S	SOCIAL ACTIONS, STATES & PROCESSES
T	TIME
W	THE WORLD & OUR ENVIRONMENT
X	PSYCHOLOGICAL ACTIONS, STATES & PROCESSES
Y	SCIENCE & TECHNOLOGY
Z	NAMES & GRAMMATICAL WORDS

Fonte: Archer et al. (2002)

Quadro 2. Subcategorias da categoria A1, 'General and abstract terms'

<b>Subcategoria de A</b>	<b>Descrição</b>
A1	General
A2	Affect
A3	Being
A4	Classification
A5	Evaluation
A6	Comparing
A7	Definite (+ modals)
A8	Seem
A9	Getting and giving; possession
A10	Open/closed; Hiding/Hidden; Finding; Showing
A11	Importance
A12	Easy/difficult
A13	Degree
A14	Exclusivizers/particularizers
A15	Safety/Danger

Fonte: Archer et al. (2002)

Quadro 3. Subcategorias da subcategoria A5, 'Education'

Subcategoria de A5	Descrição
A5.1	General
A5.1+	Evaluation: Good
A5.1-	Evaluation: Bad
A5.2	Evaluation: True/false
A5.2+	Evaluation: True
A5.2-	Evaluation: False
A5.3	Evaluation: Accuracy
A5.3+	Evaluation: Accurate
A5.3-	Evaluation: Inaccurate
A5.4	Evaluation: Authenticity
A5.4+	Evaluation: Authentic
A5.4-	Evaluation: Unauthentic

Fonte: Archer et al. (2002)

Os textos do corpus foram etiquetados automaticamente pelo USAS por meio de um script desenvolvido especialmente para esta pesquisa em linguagem Python. Depois da etiquetagem, os textos etiquetados foram processados por outros scripts desenvolvidos para a pesquisa, em linguagens Python e Shell, retornando uma matriz de dados que foi submetida à análise estatística na plataforma SAS OnDemand for Academics. Os procedimentos principais foram a análise fatorial inicial, com PROC FACTOR, que indicou a presença de quatro fatores. Em seguida, foi realizada uma outra análise fatorial rotacionada com o método PROMAX para retirar os quatro fatores. Na sequência, cada texto foi pontuado em cada dimensão por meio da função PROC SCORE. Por fim, os fatores foram interpretados qualitativamente em torno dos discursos imanentes nas músicas por meio da análise de amostras de textos com maior e menor pontuação em cada dimensão.

O segundo conjunto de dados, chamado de CoLiE acústico foi elaborado a partir do API (sigla em inglês para Interface de programação de aplicações) do serviço de músicas Spotify, que fornece um link (ID) de cada música do seu banco de dados. Os autores buscaram o link de cada música no site de desenvolvedores do spotify.com. A fim de capturar cada link, foi necessário buscar a canção pelo título ou pelo título e artista conjuntamente, verificar os resultados, clicar no título desejado, clicar em *Share* e, por fim, clicar em *Copy Song Link*. A captura manual, apesar de tediosa e demorada, se fez necessária porque as diversas tentativas de captura automática dos links não forneceram bons resultados.

Após os links terem sido capturados manualmente, os arquivos resultantes foram processados por um script, gerando uma lista. Algumas músicas coletadas do website lyrics.com não possuíam link na base do Spotify. Casos como esses foram retirados do corpus CoLiE acústico. Essas ausências da base de dados do Spotify são o motivo para o corpus CoLiE acústico não ter a mesma extensão em número de canções que o corpus CoLiE lyrics.

Em suma, o *corpus* CoLiE é composto pelo CoLiE *lyrics*, com 201.235 letras de músicas, que foram o *input* para gerar as dimensões semânticas da pesquisa; e, ainda, pelo *corpus* CoLiE acústico, formado por 97.381 músicas, cujos dados foram o *input* para as dimensões acústicas da pesquisa. As medidas acústicas obtidas pela plataforma Spotify são as seguintes:

- Musicalidade: Medida de confiança que vai de 0 a 1 e mostra se a faixa é ou não acústica. Valores próximos a 1 representam alta confiança de que a faixa é acústica.
- Dançabilidade: Descreve o potencial de uma faixa ser dançante, baseado numa combinação de elementos que incluem compasso, estabilidade do ritmo, força da batida e regularidade. Um valor de 0 é o menos dançável e 1 é o mais dançável.
- Energia: Medida perceptual da intensidade e atividade, que varia entre 0 e 1. Tipicamente, faixas energéticas são rápidas, altas e barulhentas. Características que fazem parte desta medida são loudness, timbre e entropia, de maneira geral.
- Instrumentalidade: Prevê se uma faixa contém ou não vocais. Quanto mais próximo o valor estiver de 1, maior a chance de a faixa ser instrumental. Valores acima de 0,5 tendem a representar faixas instrumentais.
- Tom/tonalidade: Valor que representa as notas predominantes. Por exemplo, 0 corresponde à escala de Dó, 1 à de Dó sustenido, 2 à de Ré etc.
- Vivacidade: Detecta a presença de público na gravação. Quanto mais alto os valores, maior a probabilidade de que a faixa seja ao vivo.
- Volume: Valor geral de uma faixa em decibéis (dB). Os valores de loudness são medidos por toda a faixa e podem comparar o volume das faixas.
- Modalidade: Indica o acorde (maior ou menor) de uma faixa. Maior é representado por 1, e menor por 0.
- Verbosidade: Detecta a presença de palavras faladas em uma faixa. Quanto maior a verbosidade, mais próximo de 1 é o valor.
- Ritmo: Estima o ritmo de uma faixa em batidas por minuto (BPM). Na terminologia musical, tempo vem a ser a velocidade de uma faixa e deriva diretamente da duração média da batida.
- Valência: Medida numérica entre 0 e 1 que descreve a positividade musical que uma faixa possui. Quanto mais alto o valor, mais positiva a música é, enquanto valores mais baixos representam músicas negativas.
- Compasso: Especifica a quantidade de batidas em cada medida, com valores de 3 a 7 (3/4, 4/4, 5/4, 6/4, 7/4), em notas de 1/4.

Uma vez as medidas acústicas terem sido coletadas, os valores foram coligidos em uma planilha, que foi submetida à análise fatorial na plataforma estatística SAS OnDemand for Academics, por meio do procedimento PROC FACTOR, que indicou a presença de três fatores. Foi então realizada uma segunda análise fatorial, rotacionada com o método PROMAX, que extraiu os três fatores. Em seguida, foram calculados os escores de fator para cada canção em cada fator, ainda no SAS OnDemand, por meio do PROC SCORE. Esses fatores foram interpretados qualitativamente por meio da escuta de uma grande amostra das canções com maior pontuação em cada fator.

Além da Análise Multidimensional de cada um dos modos (verbal e acústico) em separado, a presente pesquisa efetuou a análise estatística conjunta dos dois modos, por meio de Análise de Correlação Canônica, a qual permite o cômputo de medidas de correlação entre dois conjuntos de variáveis, em vez de apenas entre duas variáveis. Os dados para a Análise de Correlação Canônica foram os escores de cada canção em cada

dimensão de cada uma das análises multidimensionais. O cálculo foi executado no SAS OnDemand for Academics, por meio do procedimento PROC CANCOR, que retornou os pares de variantes canônicas a serem analisados. Esses pares foram interpretados qualitativamente por meio da análise textual e acústica de uma grande amostra de canções com escores destacados nas dimensões correlacionadas.

### 3. RESULTADOS

#### 3.1 Dimensões lexicais

A composição das quatro dimensões lexicais aparece no Quadro 4; os valores em parênteses indicam o peso estatístico de cada variável na solução fatorial.

Quadro 4. Padrão fatorial da análise lexical

Dimensão	Variáveis lexicais com seus respectivos pesos
Dimensão 1	Food (F1) (0,60), Clothes and personal belongings (B5) (0,52), Business (I2) (0,52), Living creatures: animals, birds, etc. (L2) (0,57), Drinks and alcohol (F2) (0,51), Sports and games generally (K5) (0,46), Money generally (I1) (0,48), The Media (Q4) (0,49), Vehicles and transport on land (M3) (0,41), Electricity and electrical equipment (O3) (0,42), Education in general (P1) (0,45), Areas around or near houses (H3) (0,39), Sailing, swimming, etc. (M4) (0,37), People (S2) (0,34), Information technology and computing (Y2) (0,35), Science and technology in general (Y1) (0,32), Architecture, houses and buildings (H1) (0,39), Drama, the theatre and show business (K4) (0,33), Linguistic Actions, States and Processes; Communication (Q1) (0,35).
Dimensão 2	Language, speech and grammar (Q3) (0,39), Psychological Actions, States and Processes (X1) (0,32), Trying (X8) (0,37).
Dimensão 3	Geographical terms (W3) (0,55), The universe (W1) (0,46), Light (W2) (0,40), Flying and aircraft (M5) (0,33), Religion and the supernatural (S9) (0,33).
Dimensão 4	Moving, coming and going (M1) (0,31), Residence (H4) (0,36).

Fonte: elaborado pelos autores

A dimensão 1 corresponde ao discurso da exaltação e/ou crítica do materialismo, do mundano, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal e do efêmero. Um exemplo é a canção *Hollywood freaks*, lançada em 1999 pelo artista Beck. Nela, o cantor faz referência contínua a produtos e objetos, exaltando e ao mesmo tempo criticando o consumismo, a opulência capitalista e o relacionamento superficial:

Exemplo 1:

*Champagne and ripple, Shamans go cripple, My sales go triple, We drop lobotomy beats, Evaporated meats, On high-tech streets, We go solo, Dance floors and talk shows, Hot dogs, (...), Be your twenty million dollar fantasy, Treat you real good, Expensive jeans*  
Em contrapartida, a dimensão 2 reflete um discurso subjetivista da existência do 'outro', relatos dos dramas pessoais e interpessoais decorrentes do sofrimento resultante da impossibilidade de conhecer a intimidade do outro, dos problemas de comunicação e entendimento mútuo e a descrição de processos mentais internos. Um exemplo é a canção *I'm trying to*, de Ashley Monroe, lançada em 2018, que relata a

problemática e a frustração decorrente da tentativa de entender outra pessoa, especialmente alguém que se ama:

Exemplo 2

*Trying to read your mind is like driving when you're blind, If only all this trying made it so, How long do you try before you let it die?, What do you need to know before you know? (...), If only all this trying made it so, I'm trying to God knows I am*

A terceira dimensão refere-se ao discurso sobrenatural, místico, metafísico, figurado, lírico, poético e transcendente. Um exemplo é a canção *I'll be over you* lançada em 2007 pelo artista Danny, que descreve um cenário da natureza como uma metáfora da melancolia que sente pela impossibilidade de ser amado.

Exemplo 3:

*When all the rivers turn to sand, When there's no sky above our land, When shadows no longer fall, When all is all, I'll be over you, When all the colors fade to grey, When there's no sun to light the day, When angels no longer fly, When tears run dry, I'll be over you.*

A última dimensão, por sua vez, captura o discurso trágico-romântico da jornada pessoal, da busca da felicidade e do amor, do sofrimento pela solidão, do reencontro, do mito do lar ideal, do amor como viagem e da busca incessante pela felicidade. A canção a seguir, lançada em 2017 e intitulada *It's a whale*, da banda Deer Tick, ilustra essa dimensão.

Exemplo 4:

*Heading nowhere with the last of my kind, I'm a martyr and I live on borrowed time and I'm fine, Hey, That's a girl, That's a boy, I'd love to see you fall in line, Fall in line.*

### 3.2 Dimensões acústicas

A composição das três dimensões acústicas aparece no Quadro 5; como anteriormente, os valores em parênteses indicam o peso estatístico de cada variável na solução fatorial.

Quadro 5. Padrão fatorial da análise acústica

Dimensão	Variáveis acústicas com seus respectivos pesos
Dimensão 1 – Polo positivo Dimensão 1- Polo negativo	energia (0,85), volume (0,79), ritmo (0,25), vivacidade (0,16). musicalidade (-0,74).
Dimensão 2 – Polo positivo Dimensão 2 - Polo negativo	dançabilidade (0,69), valência (0,65), compasso (0,24), verbosidade (0,23). instrumentalidade (-0,28).
Dimensão 3 – Polo positivo Dimensão 3 - Polo negativo	modalidade (0,38). tom (-0,29).

Fonte: elaborado pelos autores

A primeira dimensão possui dois polos, cada um capturando uma musicalidade diferente. No polo positivo, a interpretação apontou para uma musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado e voltada primeiramente para o engajamento com a plateia em performance ao vivo. Exemplos de canções bastante marcadas por esse polo são *White Riot*, de Cock Sparrer (<https://open.spotify.com/track/34iEDajZdKvoDBxPnaen1y>), *Self vs. Self*, de Pendulum

in *Flames* (<https://open.spotify.com/track/2IN6G35gsXkA3xzPYqmis5>), e *The American in me* (live) do grupo *The Avengers* (<https://open.spotify.com/track/1R0PdvKuCpgAA5GFfK6pKY>).

Já no polo negativo, a interpretação das medidas acústicas apontou para a direção oposta, ou seja, uma musicalidade com maior ênfase na instrumentação e em ritmo ameno, maior introversão e contemplação, com sensibilidade vocal, suavidade, calma e sutileza de expressão. Canções marcadas por esse polo são, por exemplo, *Outside the Wall* (Remastered version) de *Pink Floyd* (<https://open.spotify.com/track/1zeyA74WeKOpYH6wtXQwkp>) e *It's all so new to me* de Jan Savitt (<https://open.spotify.com/track/26G0R3rwgX0kE68QOABFvt>).

A segunda dimensão acústica também agrega dois polos, sendo que o primeiro reflete uma musicalidade baseada em repetição e vibração, voltada para expressão corporal, em compasso simples e com predominância de vocalização e verbosidade. Gravações com essas características musicais incluem *I'm special (Skit)* (Explicit) de Mack 10 (<https://open.spotify.com/track/7GdKNXWCtofuATCoFW38Z8>), *Wake up Mr. West* de Kanye West (<https://open.spotify.com/track/62KeM7jbmYvzT5FycdiQeS>) e *Sister surprise* de Gary Numan (<https://open.spotify.com/track/6VY5SleF8ASIIW6OLzWB1b>).

O polo negativo, por sua vez, captura uma musicalidade com características opostas, ou seja, voltada para a menor predominância vocal, maior sensibilidade, intimidade, profundidade e subjetividade. Gravações com tais características incluem *Sacrified* de *Napalm Death* (<https://open.spotify.com/track/4zjJ9fVE6ZWH5pffeEaQdK>), *Time = Cause* de *Broken Social Scene* (<https://open.spotify.com/track/5VHaHZIoFYj0VtF2eZmnhl>) e *The story of a soldier* de Ennio Morricone (<https://open.spotify.com/track/3cwREafAVNjUEt0hdrNVkt>).

A terceira e última dimensão acústica também abarca dois polos. O polo positivo indica uma musicalidade baseada em acordes maiores, tonalidade mais elevada, com sonoridade menos emotiva e menos temperamental. Canções que tipicamente possuem tal musicalidade incluem *Hello Dolly!* de Louis Armstrong (<https://open.spotify.com/track/63kd4m3VFxcJpPVvtbVNAu>), *Spiritual Cramp [Symptom Reversal]* (Remixed by Joseph Biashara) de Christian Death (<https://open.spotify.com/track/7eJ6bkzHjGEMioyehFWry>) e *Fireal* de *Deftones* (<https://open.spotify.com/track/7y0MUu7TzIA7ZyieRsNK4d>).

Já o polo negativo revela uma musicalidade com traços opostos, baseada em acordes menores e tonalidade mais baixa, ao mesmo tempo em que apresenta sonoridade mais emotiva e temperamental. Gravações bastante marcadas por essas características incluem *Give it a chance* de *Foxwarren* (<https://open.spotify.com/track/7nM1sBdMH8GHaZbQIIYQSl>), *I hate today* de *Godhead* (<https://open.spotify.com/track/3YXvwbDn5GQYwi4xncGxjh>) e *Bus driver*, de *Muddy Waters* (<https://open.spotify.com/track/6aMqOnAy0Q5EkHyjbb4VXE>).

### 3.3 Dimensões canônicas

Embora as dimensões lexicais e acústicas tenham sido obtidas em análises separadas, a fim de atingir o objetivo proposto, faz-se necessário buscar a associação entre os dois conjuntos de dimensões. Para tanto, conforme explicado, empregamos a análise de correlação canônica, tendo como dado os escores de cada canção em cada uma das dimensões lexicais e acústicas. A análise de correlação canônica identificou três correlações estatisticamente significantes (vide Tabela 1). Foi aplicado um ponto

de corte de 0,3 para detecção das dimensões integrantes de cada variável canônica. Em seguida, cada par de variantes canônicas foi interpretado como uma dimensão canônica, correspondendo a um padrão distinto de associação entre as dimensões lexicais e as acústicas.

Tabela 1. Correlações canônicas

Correlações entre as Dimensões Lexicais e suas Variáveis Canônicas			
	Variável Canônica Lexical 1	Variável Canônica Lexical 2	Variável Canônica Lexical 3
Dimensão Lexical 1	0,818	0,313	0,147
Dimensão Lexical 2	0,101	0,807	-0,235
Dimensão Lexical 3	-0,530	0,512	0,061
Dimensão Lexical 4	-0,099	0,186	0,963
Correlações entre as Dimensões Acústicas e suas Variáveis Canônicas			
Dimensão Acústica 1	0,281	0,912	0,299
Dimensão Acústica 2	0,995	-0,097	-0,009
Dimensão Acústica 3	-0,070	-0,354	0,933

Fonte: elaborado pelos autores

A primeira dimensão canônica inclui as dimensões lexicais 1 (0,81) e 3 (-0,53) bem como o polo positivo da dimensão acústica 2 (0,99), combinando assim o discurso materialista / mundano (Dim. Lexical 1) com a ausência do discurso sobrenatural (escores negativos na Dim. Lexical 3), uma sonoridade ritmada<sup>10</sup>, repetitiva e dançante<sup>11</sup> (polo positivo da Dim. Acústica 2). Um exemplo é a canção *Mornington crescent*, de Belle and Sebastian', datada de 2006 (<https://open.spotify.com/track/2jHfc8NdjF0TLyIIXPV4Rn>). Em meio a um ritmo repetitivo e dançante, a letra retrata a superficialidade presente na sociedade:

Exemplo 5:

***Kids with their spats, Ladies with chauffeurs, Dogs wearing hats and jackets, Rich apartments, Old punk posters, Tartan garments, I love the exquisite array.***

A segunda dimensão canônica, por sua vez, combina (em ordem de magnitude) as dimensões lexicais 2 (0,80), 3 (0,51) e 1 (0,31), além do polo positivo da dimensão acústica 1 (0,91) e do polo negativo da dimensão acústica 3 (-0,35); desse modo, agrega o discurso subjetivista (Dim. Lexical 2), místico / sobrenatural (Dim. Lexical 3), assim como, em menor medida, o discurso materialista / mundano (Dim. Lexical 1), juntamente com uma sonoridade intensa<sup>12</sup>, vigorosa e alta (Dim. Acústica 1), não melancólica ou temperamental (polo negativo da Dim. Acústica 3). Um exemplo é a canção *A criminal*

<sup>10</sup> A sonoridade ritmada pode ser notada quando as batidas da música (da bateria ou do sintetizador) estão mais evidentes

<sup>11</sup> A medida acústica da dançabilidade parece combinar prioritariamente elementos como estabilidade e regularidade rítmicas, produzindo altos valores mesmo em músicas com ritmo mais lento.

<sup>12</sup> Sonoridade intensa diz respeito ao timbre forte em que a música se encontra

mind, de Gowan, lançada em 1985  
(<https://open.spotify.com/track/6cwpu9CaBztS9BO2z1m80z>):

Exemplo 6:

*Before you **read** my **sentence**, I'd like to say a few **words**, (...) Some **people struggle** daily, They **struggle** with their conscience, Till the end, I have no guilt to **haunt** me*

A última correlação canônica compreende a dimensão lexical 4 (0,96) e o polo positivo da dimensão acústica 3 (0,93), refletindo, assim, canções com discurso trágico-romântico e uma sonoridade alegre<sup>13</sup>. Um exemplo é *On my way back home*, da *Band of Horses*, lançada em 2010 (<https://open.spotify.com/track/5qJw5V16ISycNIENMO1RD0>):

Exemplo 7:

*On my way back **home**, Every **step**, a victory it was, I was cheating death, just in time I woke, My memories start to **wander off**, Come to me, (...) On my way back **home***

Em resumo, as dimensões canônicas refletem três padrões linguístico-musicais de larga escala:

1. Discurso materialista mundano com sonoridade ritmada;
2. Discurso subjetivo místico com sonoridade intensa;
3. Discurso trágico romântico com sonoridade alegre;

#### 4. DISCUSSÃO

Os achados da pesquisa dialogam com a literatura prévia de três maneiras principais. A primeira é revisitando a noção de que as letras de música popular em inglês são vazias e superficiais, resumindo-se à temática do amor e sexo (COLE, 1971; PEATMAN, 1944; cf. SHUKER, 2001, p. 146). Subjacente à aparente banalidade, há na verdade discursos que expressam uma determinada posição do sujeito e visão do mundo.

Em segundo lugar, os achados mostram que essa complexidade discursiva das letras de música popular não se restringe a compositores consagrados. Segundo SHUKER (2001, p. 146-147), compositores de rock foram alçados a um patamar de prestígio nos anos 1960, transformando esse gênero musical em uma forma artística de alto valor cultural. Assim, letras de Bob Dylan, por exemplo, passaram a ser vistas como poesia musical, contrapondo-se às letras de outros compositores, considerados de pouco valor estético. Embora esta pesquisa não tenha focado o estilo de compositores, tampouco a qualidade lírica da composição, dada a alta frequência de discursos representativos do mundo contemporâneo nas letras, é possível sugerir que esses discursos não refletem o estilo de compositores individuais.

Por fim, a presente pesquisa mostra que em um nível macro, existe uma relação de codependência entre o texto e os elementos fundamentais da canção (o ritmo, a melodia e a harmonia), no sentido em que determinadas letras tendem a preferir determinadas sonoridades e vice-versa. Esse achado vai ao encontro de posições como a de North, Krause e Ritchie (2020, p. 3), segundo os quais *os músicos estão sujeitos a um imperativo artístico que asseguram que a música e qualquer letra se alinham de uma certa maneira uma com a outra para facilitar a comunicação*<sup>14</sup>. Os autores citam os

<sup>13</sup> Sonoridade alegre se baseia no fato de a música ter sido composta em tom maior (dó maior, por exemplo).

<sup>14</sup> No original: musicians are subject to an artistic imperative to ensure that music and any lyrics in some way align with one another in order to facilitate communication.

Beatles como exemplo, pois no início da carreira da banda, John Lennon e Paul McCartney deliberadamente incluíam pronomes pessoais de primeira pessoa nas letras e compunham estruturas musicais simples a fim de maximizar o engajamento com o público. Os padrões identificados dificilmente são deliberados, isto é, planejados conscientemente pelos compositores. As dimensões mostram uma gradação da aderência das composições aos discursos subjacentes, com determinadas canções exibindo grande concentração e outras baixa concentração ou total ausência dos discursos.

## 5. COMENTÁRIOS FINAIS

Este trabalho apresentou uma descrição da música popular anglo-saxã baseada em corpus envolvendo tanto o modo verbal quanto o modo acústico. A partir de descrições detalhadas baseadas em grandes quantidades de dados e em procedimentos estatísticos multivariados, foi possível detectar padrões linguístico-discursivos nas letras de música bem como padrões acústicos na performance musical. Além disso, foi possível detectar três macro-padrões multimodais de escolhas linguísticas e escolhas acústicas. Os resultados apontam, de modo geral, para uma padronização complexa das letras de música de um lado, das sonoridades de outro, e da simbiose entre texto e melodia de outro.

A pesquisa aqui relatada pretende contribuir para a Análise Multidimensional lexical de dois modos. Primeiramente, na medida em que emprega a etiquetagem semântica como meio de processar o léxico, submetendo-o à uma taxonomia classificativa. Nesse sentido, os resultados mostraram que as categorias semânticas permitem uma visão hierárquica dos sentidos expressos pelo vocabulário das músicas, o que seria mais difícil de obter sem a etiquetagem semântica. E em segundo lugar, porque demonstra ser possível descrever registros levando em conta não somente o modo verbal, mas também o acústico. Isso foi possível devido à tecnologia de processamento de som empregada pela plataforma Spotify, por meio da qual são produzidos índices acústicos das composições musicais. Este trabalho vem se juntar a outros que empregam ferramentas computacionais baseadas em Inteligência Artificial para processamento de grandes quantidades de imagens (BERBER SARDINHA, 2021b), inaugurando a vertente multimodal da Análise Multidimensional.

---

## REFERÊNCIAS

- BAKER, PAUL; MCENERY, TONY. Introduction. In: BAKER, PAUL e MCENERY, TONY. **Corpora and Discourse Studies: Integrating Discourse and Corpora**. Basingstoke: Palgrave Macmillan. p. 1-20. 2015.
- BARBOSA, F. Um corpus do samba carioca para estudos lexicográficos e discursivos. In: SHEPHERD, TANIA; BERBER SARDINHA, TONY e VEIRANO PINTO, MARCIA. **Caminhos da Linguística de Corpus**. Campinas, SP: Mercado de Letras. p. 247-270. 2012.
- BENNETT, ANDY. Towards a cultural sociology of popular music. **Journal of Sociology**, 44, n. 4, p. 419-432, 2008.
- BERBER SARDINHA, TONY. Linguística de Corpus: histórico e problemática: **DELTA**, 16, n.2, 2000.
- BERBER SARDINHA, TONY. **Linguística de Corpus**. São Paulo: Manole, 410pp., 2004.
- BERBER SARDINHA, TONY; VEIRANO PINTO, MARCIA (ed.). **Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber**. Amsterdam/Philadelphia, PA: John Benjamins, 2014.
- BERBER SARDINHA, TONY. ANÁLISE MULTIDIMENSIONAL. **DELTA: Documentação E Estudos**  
Em Linguística Teórica E Aplicada, 16(1).  
<https://revistas.pucsp.br/index.php/delta/article/view/39990>, 2018.

- BERBER SARDINHA, TONY. Using multi-dimensional analysis to detect representations of national culture. In T. Berber Sardinha & M. Veirano Pinto (Eds.), **Multi-dimensional analysis: Research methods and current issues** (pp. 231-258). Bloomsbury / Continuum, 2019.
- BERBER SARDINHA, TONY. A historical characterization of American and Brazilian cultures based on lexical representations. In: **Corpora**, v.14, n.2, pp. 183-212, 2020.
- BERBER SARDINHA, TONY. Discourse of academia from a multi-dimensional perspective. In E. Friginal & J. Hardy (Eds.), **The Routledge Handbook of Corpus Approaches to Discourse Analysis** (pp. 298-318). Routledge, 2021a.
- BERBER SARDINHA, TONY. Going multimodal in corpus linguistics: The case of social media. **International perspectives on corpus technology for language learning**, University of Queensland, 2021b.
- BERBER SARDINHA, TONY; KAUFFMANN, CARLOS. Brazilian Portuguese literary style. In FRIGINAL, ERICK & HARDY, JACK (Eds.), **The Routledge handbook of corpus**. Abingdon: Routledge, 2021.
- BERBER SARDINHA, TONY; ALENCAR, ANDRÉ L.S.; DA SILVA, CICERO S.; GIL, CRISTINA B.; LOPES, MAURÍCIO J.F. HUGHES, STEPHAN A.S. “#EUNAOVOUTOMARVACINA”: Uma abordagem da linguística de corpus e da análise multimodal imagética. **Revista Intercâmbio**, v.LI: 178-209, 2022. São Paulo: LAEL/PUCSP. ISSN 2237-759X 178 <https://doi.org/10.23925/2237.759X.2022V51.e58515>, 2022.
- BERBER SARDINHA, TONY. Corpus linguistics and historiography: Finding the major discourses in the first 50 years of tesol quarterly. **Journal of Research Design and Statistics in Linguistics and Communication Science**, 7(1), 69-90, 2023.
- BÉRTOLI, PATRÍCIA. **Explorando a Linguística de corpus e letras de música na produção de atividades pedagógicas**. MA Thesis (Dissertação de mestrado) - LAEL, PUCSP, São Paulo, SP. 2002.
- BÉRTOLI, PATRÍCIA. **Linguagem da música popular anglo-americana de 1940-2009**. PhD Dissertation (Tese de Doutorado) - LAEL, PUCSP, São Paulo, SP. 2010.
- BIBER, DOUGLAS. On the investigation of spoken/written differences. **Studia Linguistica**, 40, p. 1-21, 1986.
- BIBER, DOUGLAS. **Variation across speech and writing**. Cambridge: Cambridge University Press. 1988.
- BIBER, DOUGLAS. An analytical framework for register studies. In: BIBER, D. e FINEGAN, E. **Sociolinguistic perspectives on register**. Oxford: Oxford University Press. p. 31-56. 1994.
- BIBER, DOUGLAS. **Dimensions of Register Variation - A Cross-Linguistic Comparison**. Cambridge: Cambridge University Press. 1995.
- BIBER, DOUGLAS. Conversation text types: A multi-dimensional analysis. In: PURNELLE, GÉRALD ;FAIRON, CÉDRICK e DISTER, ANNE. **Le poids des mots: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data**. Louvain: Presses universitaires de Louvain. p. 15-34. 2004.
- BIBER, DOUGLAS. Corpus-based analyses of discourse: Dimensions of variation in conversation. In: BHATIA, V.;FLOWERDEW, J. e JONES, R. **Advances in Discourse Studies**. London: Routledge. p. 110-114. 2008.
- BIBER, DOUGLAS. Register as a predictor of linguistic variation. **Corpus Linguistics and Linguistic Theory**, 8, n. 1, p. 9-37, 2012.
- BIBER, DOUGLAS. Text-linguistic approaches to register variation. **Register Studies**, 1, n. 1, p. 42-75, 2019.
- BIBER, DOUGLAS; CONRAD, S.; REPPEN, R. **Corpus Linguistics - Investigating Language Structure and Use**. Cambridge: Cambridge University Press. 1998.
- BIBER, DOUGLAS; CONRAD, SUSAN. **Register, genre, and style**. Cambridge ; New York: Cambridge University Press. ix, 344 p. 2009.
- BURR, V. **An Introduction to Social Constructionism**. London: Routledge. 1995.
- CLARKE, ISOBELLE; MCENERY, TONY; BROOKES, G. Multiple Correspondence Analysis, newspaper discourse and subregister: A case study of discourses of Islam in the British press. **Register Studies**, 3, n. 1, p. 144-171, 2021.
- COLE, RICHARD R. Top songs in the sixties -- A content analysis of popular lyrics. **American Behavioral Scientist**, 14, n. 3, p. 389-400, 1971.
- DELFINO, MARIA CLAUDIA N. Uso de Música para o Ensino de Inglês como Língua Estrangeira em um Ambiente Baseado em Corpus [MA Thesis, PUCSP], 2016.
- DELFINO, MARIA CLAUDIA N., ARAÚJO, RAFAEL F., & BERBER SARDINHA, TONY. Revista Brasileira de Linguística Aplicada: multidimensões temáticas. In M. J. B. Finatoo, R. R. Rebechi,

- S. Sarmento, & A. E. P. Bocorny (Eds.), **Linguística de corpus : perspectivas** (pp. 93-125). Editora do Instituto de Letras – UFRGS, 2018.
- DUNBAR-HALL, P. Semiotics as a method for the study of popular music. **International Review of the Aesthetics and Sociology of Music**, 22, n. 2, p. 127-132, 1991.
- FITZSIMMONS-DOOLAN, SHANNON. Using lexical variables to identify language ideologies in a policy corpus. **Corpora**, 9, n. 1, p. 57-82, 2014.
- FITZSIMMONS-DOOLAN, SHANNON. Language ideologies of institutional language policy: Exploring variability by language policy register. **Language Policy**, 18, n. 2, p. 169-189, 2019.
- GRAY, BETHANY. **Linguistic variation in research articles: When discipline tells only part of the story**. Amsterdam: John Benjamins. 2015.
- HAJER, M. Discourse coalitions and the institutionalization of practice. *In: The argumentative turn in policy analysis and planning* Durham, NC: Duke University Press. p. 43-76. 1993.
- HALL, STUART. **Formations of Modernity**. Cambridge: Polity Press. 1992.
- HARDY, JACK A. Multi-Dimensional analysis of academic discourse. *In: BAKER, PAUL e MCENERY, TONY. Corpora and Discourse Studies: Integrating Discourse and Corpora*. Basingstoke: Palgrave Macmillan. p. 155-174. 2015.
- HOEY, MICHAEL. **Lexical Priming: A New Theory of Words and Language**. London, New York: Routledge. xiii, 202 p. p. 2005.
- HUNSTON, SUSAN; FRANCIS, GILL. Pattern grammar : a corpus-driven approach to the lexical grammar of English. **Studies in corpus linguistics**, n. 4, p. xi, 288 p., 2000.
- ANONIMIZADO. 2021.
- MCENERY, TONY; WILSON, ANDREW. **Corpus linguistics**. Edinburgh: Edinburgh University Press. 209 p. p. 1996.
- MIDDLETON, R. Popular music analysis and Musicology: Bridging the gap. **Popular Music**, 12, n. 2, p. 177-190, 1993.
- MILANEZ, ALINE ZAMBONI; KAUFFMANN, CARLOS; DELFINO, MARIA CLAUDIA N.; BERTOLI DUTRA, PATRICIA *et al.* As dimensões do twitter contra e a favor das manifestações do 7 de setembro de 2021. *In: 22º InPLA – Intercâmbio de Pesquisas em Linguística Aplicada*, Sao Paulo, PUCSP. 2021.
- MOORE, ALLAN F. **Song means: Analysing and interpreting recorded popular song**. Farnham: Ashgate. 2021.
- NORTH, ADRIAN C; KRAUSE, AMANDA E; RITCHIE, DAVID. The relationship between pop music and lyrics: A computerized content analysis of the United Kingdom's weekly top five singles, 1999–2013. **Psychology of Music**, Ahead of print, p. 1-24, 2020.
- PEATMAN, JOHN GRAY. Radio and popular music. *In: LAZARSELD, PAUL F. e STANTON, FRANK N. Radio Research 1942-1943*. New York: Essential Books. p. 335-396. 1944.
- ROMEIRO, YARA. **A linguagem verbal das artes visuais: uma análise multidimensional do discurso sobre a fotografia de Sally Mann**. (Dissertação de mestrado) - LAEL, PUCSP, São Paulo. 2020.
- SHUKER, R. **Understanding popular music**. 2nd ed. London: Routledge. 2001.
- SINCLAIR, JOHN MCH. Beginning the study of lexis. *In: BAZELL, C. E. In Memory of J R Firth*. London: Longman. p. 410-430. 1966.
- SINCLAIR, JOHN MCH. Naturalness in language. **ELR Journal**, 2, p. 11-20, 1988.
- SINCLAIR, JOHN MCH. **Corpus, Concordance, Collocation**. Oxford, New York: Oxford University Press. xviii, 179 p. p. 1991.
- SINCLAIR, JOHN MCH.; JONES, SUSAN; DALEY, ROBERT. **English Lexical Studies: The OSTI Report**. London/New York: Continuum. 2-138 p. 1970/2004.
- VEIGA, ALEXANDRE TRIGO. **As dimensões da fé: Sete religiões mundiais em análise multidimensional lexical**. Tese de doutorado -, LAEL, PUCSP, São Paulo. 2021.
- VLADI, N. O negócio da música – como os gêneros musicais articulam estratégias de comunicação para o consumo cultural. *In: GP Comunicação e Culturas Urbanas (DT6), X Encontro dos Grupos de Pesquisas em Comunicação, evento componente do XXXIII Congresso Brasileiro de Ciências da Comunicação, Caxias do Sul, RS. 2010.*
- WALL, TIM. **Studying Popular Music Culture**. Newbury Park: Sage. 2013.
- WERNER, VALENTIN. Love is all around: A corpus-based study of pop music lyrics. **Corpora**, 7, n. 1, p. 19-50, 2012.
- WERNER, VALENTIN. Catchy and conversational? A register analysis of pop lyrics. **Corpora**, 16, n. 2, p. 237-270, 2021.
- ZUPPARDI, MARIA CAROLINA, & BERBER SARDINHA, TONY. A multi-dimensional view of collocations in academic writing. *In U. Römer, V. Cortes, & E. Friginal (Eds.), Advances in*

**Corpus-based Research on Academic Writing. Effects of Discipline, Register, and Writer Expertise** (pp. 334–353). John Benjamins 2020.

Recebido: 21/12/2022

Aceito: 29/7/2023

Publicado: 4/12/2023