

Pontifícia Universidade Católica de São Paulo

PUC-SP

Maria Claudia Nunes Delfino

***More than words: análise multimodal multidimensional da música
popular em língua inglesa***

Doutorado em Linguística Aplicada e Estudos da Linguagem

SÃO PAULO

2022

Pontifícia Universidade Católica de São Paulo

PUC-SP

Maria Claudia Nunes Delfino

***More than words: análise multimodal multidimensional da música
popular em língua inglesa***

Doutorado em Linguística Aplicada e Estudos da Linguagem

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

SÃO PAULO

2022

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta tese de doutorado por processos fotocopiadores ou eletrônicos.

Assinatura: *Ma. Claudia N. Delfino*

Data: 30/08/2022

e-mail: claudia@fatecpg.com.br

Currículo Lattes: <http://lattes.cnpq.br/6401133997423273>

Delfino, Maria Claudia Nunes

More than words: análise multimodal multidimensional da música popular em língua inglesa / Maria Claudia Nunes Delfino. São Paulo: 2022

xiv + 141p.

Orientador: Professor Doutor Antonio Paulo Berber Sardinha

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.

Área de concentração: Linguística Aplicada e Estudos da Linguagem

1. Análise Multidimensional. 2. Música. 3. Big Data. 4. Linguística de Corpus. 5. Multimodalidade

Maria Claudia Nunes Delfino

More than words: análise multimodal multidimensional da música popular em língua inglesa.

Aprovado em / / .

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

Banca Examinadora

Para meus pais, **Elinaldo** e **Lúcia**, pelo amor incondicional.

Para **Tony**, mais que o orientador do trabalho e da vida, por ser a pessoa que acreditou em mim quando nem eu acreditava.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Programas: PROSUP / PROSUC

Número do Processo: 88887.177403/2018-00

Período: 01/03/2018 a 31/08/2022

Instituição: Pontifícia Universidade Católica de São Paulo

Programa: PDSE

Número do Processo: 88881.362173/2019-01

Período: 01/08/2019 a 31/01/2020

Instituição: Northern Arizona University (Flagstaff, EUA)

AGRADECIMENTOS

Agradeço, em primeiro lugar, ao meu orientador, Prof. Dr. Tony Berber Sardinha, por representar o pesquisador e a pessoa que um dia, depois de muito estudar a língua em uso, espero me tornar. Faço um agradecimento especial ao Prof. Dr. Jesse Egbert pela orientação da pesquisa durante meu período de estudos na Northern Arizona University e ao Prof. Dr. Joe Collentine pela ajuda no desenvolvimento dos scripts para esta pesquisa.

Gostaria de agradecer aos Profs. Drs. Cristina Mayer, Jesse Egbert, Douglas Biber, Randi Reppen, Tove Larsson, Joe Collentine, Patrícia Bértoli e Ana Elisa Machado Cysne, pela leitura atenta das versões que resultaram nesse trabalho. Obrigada por todas as sugestões, essenciais na redação final desta tese.

Agradeço a todos os pesquisadores do Gelc pelas críticas, ideias, sugestões, ajuda com a parte acústica do trabalho. Vocês foram essenciais para essa pesquisa tomar corpo. Sem esgotar a lista, aqui vão alguns nomes que sempre estiveram presentes: Márcia Veirano Pinto, Simone Vieira Resende, Rafael Fonseca de Araújo, Maria Carolina Zuppari, Alexandre Trigo Veiga, Yara Toledo, Aline Andrea Zamboni Milanez, Carlos Henrique Kauffmann, Cicero Soares da Silva, Cristina Borges Gil, Eduardo Batista da Silva, Paula Tavares Pinto, Katherine Oliva, Luiz Carlos Zeferino, Marcos Oliveira, Renata Sant Anna Lamberti Spagnuolo e Stephan Hughes.

Agradeço à equipe do LAEL pela organização acadêmica, e à dedicação de Maria Lúcia dos Reis e Marcia Martins, do CEPRIL. Agradeço à Profa. Dra. Maximina Maria Freire por oferecer um novo olhar sobre o ensinar.

Agradeço a minha querida família por todo o carinho demonstrado em todas as horas, especialmente meus pais, irmãos, sobrinhos, minha família de Ribeirão Preto, com especial carinho a Márcia e o Fábio. Estendo o agradecimento aos meus amigos, meus queridos companheiros e alunos da Fatec Praia Grande, com especial carinho ao Ulysses Diegues e a Tatiana Schmitz, ao meu super secretário de mocidade (Rafa) e a melhor equipe de coordenação (Alice e Gabriel) que poderia ter pedido na vida.

I'd also like to thank some very important friends I made in Flagstaff during the time I lived there: Klaire and Richard for sharing a home with me, Ana Bocorny and Larissa Goulat, my Brazilian friends who were always there for me. Joe and Karina Collentine, my "American Family", Tove Larsson, with whom I shared the best pizzas and laughter on Sunday evenings and the graduate students from the Applied Linguistics Department at NAU during the fall semester of 2019.

Por fim, mas não menos importante, agradeço ao leitor e ao Carlos Kauffmann pela revisão.

Alex Fletcher : It doesn't have to be perfect. Just spit it out. They're just lyrics.

Sophie Fisher : "Just lyrics"?

Alex Fletcher : Lyrics are important. They're just not as important as melody.

Sophie Fisher : I really don't think you get it. A melody is like seeing someone for the first time. The physical attraction. Sex.

Alex Fletcher : I so get that.

Sophie Fisher : But then, as you get to know the person, that's the lyrics. Their story. Who they are underneath. It's the combination of the two that makes it magical.

(Filme: Letra e música, 2007)

RESUMO

A música popular tem sido objeto de diversos estudos nas mais variadas áreas, como a Musicologia (MIDDLETON, 1993; MOORE, 2021), Estudos Culturais (WALL, 2013), Sociologia (BENNETT, 2008), Semiótica (DUNBAR-HALL, 1991), Linguística (BARBOSA, 2012; BÉRTOLI DUTRA, 2010; BERTOLI-DUTRA, 2014; DELFINO, 2016; WERNER, 2012), entre outras. Embora exista grande interesse no estudo da música popular, há ao menos quatro tendências gerais na literatura relevantes à presente tese. A primeira é o número reduzido de pesquisas empíricas sobre a linguagem musical, visto que há uma predominância de trabalhos de natureza teórica. A segunda é o reducionismo, ou seja, a desvinculação do componente verbal (a letra das músicas) do componente acústico (a melodia). A terceira tendência, entre os estudos empíricos, diz respeito ao emprego 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. Por fim, a quarta tendência é a correlação entre o modo verbal e o modo acústico da música. Há, portanto, quatro lacunas significativas na literatura prévia, uma relativa à escassez de estudos empíricos, uma segunda que diz respeito à escassez de estudos envolvendo tanto o componente verbal quanto o acústico. A terceira concerne à escassez de estudos baseados em grandes amostras de dados e a última referente à falta de uma correlação musical que leve em conta as características linguísticas das letras de música e as características acústicas da produção musical simultaneamente. Tendo em vista essas lacunas, o presente trabalho tem como objetivo realizar duas análises multidimensionais, quais sejam: uma semântica e outra acústica, outro objetivo é observar a correlação entre o modo verbal e o modo acústico por meio de uma análise de correlação canônica, um procedimento estatístico que relaciona conjuntos de mensurações, no caso os escores de cada canção em cada uma das dimensões semânticas e acústicas. Essa correlação possibilitou a extração de três dimensões multimodais: Dimensão 1 - Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade, retratando o discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero distanciando-se do sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente (Musicalidade para dançar retratando o discurso

materialista concreto e distanciando-se do místico); Dimensão musical 2 - Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo, com 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, descrição de processos mentais internos, descrevendo o mundo como sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente (Musicalidade intensa retratando o discurso do drama pessoal e do sobrenatural); Dimensão musical 3 - Musicalidade baseada em acordes mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental retratando 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, da busca incessante pela felicidade (Musicalidade alegre retratando o discurso trágico-romântico).

Palavras-chave: Análise Multidimensional; Música; Big Data, Linguística de Corpus; Multimodalidade

ABSTRACT

Popular music has been the object of several studies in various fields, such as Musicology (MIDDLETON, 1993; MOORE, 2021), Cultural Studies (WALL, 2013), Sociology (BENNETT, 2008), Semiotics (DUNBAR-HALL, 1991), and Corpus Linguistics (BARBOSA, 2012; BÉRTOLI DUTRA, 2010; BERTOLI-DUTRA, 2014; DELFINO, 2016; WERNER, 2012), among others. While there is great interest in the study of popular music, there are at least four general trends in the literature relevant to the present thesis. The first is the small number of empirical research on musical language, as there is a predominance of works of a theoretical nature. The second is reductionism, that is, the disconnection of the verbal component (the lyrics of songs) from the acoustic component (the melody). The third trend, among empirical studies, concerns the use of reduced data samples, formed by a few songs, which reduces the scope of the study and therefore the possibility of generalizing the results. Finally, the fourth trend is the correlation between the verbal and the acoustic modes of the songs. There are therefore four significant gaps in the previous literature, one concerning the scarcity of empirical studies, a second concerning the scarcity of studies involving both the verbal and acoustic components. The third concerning the scarcity of studies based on large data samples, and the last concerning the lack of a musical correlation that takes into account the linguistic characteristics of lyrics and the acoustic characteristics of musical production simultaneously. In view of these gaps, the present thesis aims at performing two multidimensional analyses, namely: a semantic analysis (verbal mode) and an acoustic analysis (acoustic mode), so that each song is measured along several dimensions simultaneously; another goal is to compute the correlation between the verbal and the acoustic modes through a canonical correlation analysis, a statistical procedure that relates canonical pairs of measurement sets, in this case the scores of each song on each of the semantic and acoustic dimensions. This correlation enabled the extraction of three multimodal musical dimensions: Dimension 1 - Musicality based on repetitive, vibrant rhythm, focused on bodily expression indoors, in conventional tempo, with a predominance of vocalization and verbosity, portraying the discourse of glorifying and/or criticizing materialism, wealth, luxury, ostentation futility, opulence, instant gratification, superficiality, appearance, the banal, the ephemeral, distancing itself from the supernatural, mystical, metaphysical, figurative, lyrical, poetic, transcendent (Musicality for dancing portraying the concrete materialist discourse and distancing itself from the

mystical); Musical Dimension 2 - Musicality based on intensity, amplified instrumentation, extroversion, vigorous, powerful vocalization, with high volume, strength and fast movement, in accelerated rhythm, aimed primarily at engaging with the audience in live performance, with subjectivist discourse on the existence of the 'other', accounts of personal and interpersonal dramas resulting from suffering due to the impossibility of knowing the intimacy of the other, problems of communication and mutual understanding, description of internal mental processes, describing the world as supernatural, mystical, metaphysical, figurative, lyrical, poetic, transcendent (Intense musicality depicting the discourse of personal drama and the supernatural); Musical Dimension 3 - Musicality based on happier chords and higher tones, with a less emotional and less moody sound portraying the tragic-romantic discourse of the personal journey, the search for happiness and love, suffering through loneliness, reunion, the myth of the ideal home, love as a journey, the ceaseless search for happiness (Joyful musicality portraying the tragic-romantic discourse).

Keywords: Multidimensional analysis; Music; Big Data; Corpus Linguistics; Multimodality

SUMÁRIO

1. INTRODUÇÃO	1
2. FUNDAMENTAÇÃO TEÓRICA	5
2.1 LINGUÍSTICA DE <i>CORPUS</i>	6
2.1.1 <i>Análise Multidimensional</i>	8
2.1.2 <i>Anotação semântica</i>	11
2.1.3 <i>Multimodalidade: o acústico e o verbal na música</i>	15
2.2 ANÁLISE DE CORRELAÇÃO CANÔNICA	19
2.3 CONSIDERAÇÕES SOBRE A MÚSICA POPULAR EM LÍNGUA INGLESA	21
3. METODOLOGIA	30
3.1 CORPORA EMPREGADOS.....	30
3.1.1 <i>Colie lyrics</i>	30
3.1.2 <i>ColiE acústico</i>	34
3.2 CARACTERÍSTICAS SEMÂNTICAS.....	38
3.3 ANOTAÇÃO SEMÂNTICA	43
3.4 ANÁLISE MULTIDIMENSIONAL SEMÂNTICA.....	44
3.5 CARACTERÍSTICAS ACÚSTICAS	50
3.6 ANÁLISE MULTIDIMENSIONAL ACÚSTICA	51
3.7 ANÁLISE DE CORRELAÇÃO CANÔNICA	53
4. RESULTADOS E DISCUSSÃO	57
4.1 RESULTADOS DA AMD SEMÂNTICA	57
4.1.1 - <i>Fator 1, análise semântica</i>	58
4.1.2 - <i>Fator 2, análise semântica</i>	63
4.1.3 - <i>Fator 3, análise semântica</i>	67
4.1.4 - <i>Fator 4, análise semântica</i>	71
4.2 RESULTADOS DA AMD ACÚSTICA.....	77
4.2.1 - <i>Fator 1, análise acústica</i>	78
4.2.2 - <i>Fator 2, análise acústica</i>	81
4.2.3 - <i>Fator 3, análise acústica</i>	85
4.3 ANÁLISE DE CORRELAÇÃO CANÔNICA	88
4.3.1 - <i>Primeiro par de variantes canônicas</i>	90
4.3.2 - <i>Segundo par de variantes canônicas</i>	93
4.3.3 - <i>Terceiro par de variantes canônicas</i>	96
5. CONSIDERAÇÕES FINAIS	100
6. REFERÊNCIAS	104
7. ANEXOS	114
ANEXO 1 - PROGRAMA SAS PARA ANÁLISE DE <i>CLUSTERS</i> DAS VARIÁVEIS	114
ANEXO 2 - PROGRAMA SAS PARA ANÁLISE FATORIAL SEMÂNTICA	117
ANEXO 3 - PROGRAMA SAS PARA ANÁLISE FATORIAL ACÚSTICA	133
ANEXO 4 - PROGRAMA SAS PARA ANÁLISE CANÔNICA.....	140

LISTA DE TABELAS

TABELA 1 - COMPOSIÇÃO DO CORPUS COLIÉ <i>LYRICS</i>	33
TABELA 2 - PADRÃO FATORIAL DA ANÁLISE SEMÂNTICA	49
TABELA 3 - VARIÁVEIS ACÚSTICAS DO SPOTIFY.....	50
TABELA 4 - PADRÃO FATORIAL DA ANÁLISE ACÚSTICA	52
TABELA 5 - TESTE DA HIPÓTESE NULA H_0	55
TABELA 6 - TESTES COMPLEMENTARES PARA ATESTAR A CONFIABILIDADE DA ANÁLISE DE CORRELAÇÃO CANÔNICA	55
TABELA 2 - PADRÃO FATORIAL DA ANÁLISE SEMÂNTICA	57
TABELA 7 – CORRELAÇÃO ENTRE AS DIMENSÕES E O TOTAL DE PALAVRAS DOS TEXTOS DO <i>CORPUS</i>	76
TABELA 4 - PADRÃO FATORIAL DA ANÁLISE ACÚSTICA	77
TABELA 8 - CANÇÕES COM ALTO VALOR DE ENERGIA	78
TABELA 9 - CANÇÕES COM ALTO VALOR DE BATIDAS POR MINUTO.....	79
TABELA 10 - CANÇÃO COM ALTO VALOR DE VOLUME.....	79
TABELA 11 - CANÇÕES COM ALTO VALOR DE VIVACIDADE	80
TABELA 12 - CANÇÕES COM ALTO VALOR DE ACUSTICIDADE (INSTRUMENTAÇÃO)	80
TABELA 13 - CANÇÕES COM ALTO VALOR DE COMPASSO	82
TABELA 14 - CANÇÕES COM ALTO VALOR DE VERBOSIDADE	82
TABELA 15 - CANÇÃO COM ALTO VALOR DE VALÊNCIA	83
TABELA 16 - CANÇÕES COM ALTO VALOR DE DANÇABILIDADE.....	83
TABELA 17 - CANÇÕES COM ALTO VALOR DE INSTRUMENTALIDADE	84
TABELA 18 - CANÇÕES COM VALOR MÁXIMO (1) DE MODALIDADE.....	85
TABELA 19 - CANÇÕES COM ALTO VALOR DE TOM.....	86
TABELA 20 - CORRELAÇÃO ENTRE AS DIMENSÕES ACÚSTICAS	88
TABELA 21 - CORRELAÇÃO CANÔNICA ENTRE AS DIMENSÕES SEMÂNTICAS E ACÚSTICAS	89
TABELA 22 - ESTRUTURA CANÔNICA.....	89
TABELA 23 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO PRIMEIRO PAR DE VARIANTES CANÔNICAS....	90
TABELA 24 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO SEGUNDO PAR DE VARIANTES CANÔNICAS ...	93
TABELA 25 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO TERCEIRO PAR DE VARIANTES CANÔNICAS	96

LISTA DE FIGURAS

FIGURA 1 – PASSOS DE UMA ANÁLISE MULTIDIMENSIONAL	9
FIGURA 2 - CAMPOS SEMÂNTICOS DO TAGSET DO ETIQUETADOR USAS	13
FIGURA 3 - CATEGORIAS E SUBCATEGORIAS DO CAMPO SEMÂNTICO TIME NO USAS	14
FIGURA 4 - TEXTO ETIQUETADO COM O USAS.....	15
FIGURA 5 - PRIMEIRA IMAGEM QUE A MTV TRANSMITIU	27
FIGURA 6 - EXEMPLO DE UM ARQUIVO COM LETRA DE MÚSICA.....	31
FIGURA 7- EXEMPLO DE UM ARQUIVO COM METADADOS	31
FIGURA 8 - PARTE DO ARQUIVO ÍNDICE	32
FIGURA 9- PARTE DO ARQUIVO ÍNDICE CONTENDO ARQUIVOS DE REGRAVAÇÕES DA MESMA MÚSICA.....	33
FIGURA 10 - COLETA DO LINK DE MÚSICAS NA PLATAFORMA SPOTIFY	35
FIGURA 11 - ARQUIVO CONTENDO OS IDS DE CADA MÚSICA DA PLATAFORMA SPOTIFY.....	36
FIGURA 12 - OBTENDO MÉTRICAS ACÚSTICAS DO SPOTIFY	36
FIGURA 13 - CATEGORIAS GERADAS PELO API DO SPOTIFY	37
FIGURA 14 - CATEGORIAS E SUBCATEGORIAS DO ETIQUETADOR USAS.....	38
FIGURA 15 - PLANILHA COM OS CAMPOS SEMÂNTICOS DO CORPUS COLIE.....	44
FIGURA 16 - GRÁFICO DE ESCARPA DA ANÁLISE SEMÂNTICA	48
FIGURA 17 - GRÁFICO DE ESCARPA DA ANÁLISE ACÚSTICA	51
FIGURA 18 - POLOS DO FATOR 1 DA ANÁLISE MULTIDIMENSIONAL ACÚSTICA	53
FIGURA 19 – SÍNTESE DA ANÁLISE MULTIMODAL PRETENDIDA NO TRABALHO	56
FIGURA 20 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 1	62
FIGURA 21 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 2	66
FIGURA 22- DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 3	71
FIGURA 23 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 4	74
FIGURA 24- ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 1	81
FIGURA 25 - ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 2.....	84
FIGURA 26 - ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 3.....	86

1. INTRODUÇÃO

(I can't get no) satisfaction¹.

Keith Richards / Mick Jagger

A música é considerada um sistema universal de sinais sonoros, humanos, dinâmicos, com múltiplos propósitos (WILLIAMSON, 2014) e que também vem a ser uma expressão artística, que pode representar a cultura em que está inserida. Em muitos casos, o gosto musical une as pessoas, dando uma sensação de pertencimento ao grupo; em outros, afasta. Grupos que gostam de um determinado estilo musical tendem a não se misturar a grupos de um estilo distinto do seu. Podemos dizer que esse pertencimento expresso através da música ajuda a moldar nossa identidade. Dada a importância de tal constructo cultural, a presente pesquisa enfocou sua constituição textual e sonora em língua inglesa. A produção musical em inglês influencia não apenas falantes nativos do inglês, mas, devido à globalização e ao domínio da indústria cultural por empresas norte-americanas e britânicas, a canção em inglês influencia o mundo inteiro.

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-DUTRA, 2010; BÉRTOLI-DUTRA, 2014; DELFINO, 2016; WERNER, 2012), entre outras. Ela não apenas é uma manifestação cultural que gera bilhões de dólares para produtores, agentes, cantores, músicos, bandas, fabricantes de equipamentos, distribuidores, mídia e outras atividades ligadas a essa indústria, como também é um produto que contribui para moldar o comportamento e a cultura de várias gerações (VLADI, 2010).

Embora exista grande interesse no estudo da música popular, há pelo menos quatro tendências gerais na literatura relevantes à presente tese. 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

¹ “Eu não consigo nenhuma satisfação” (Música de mesmo título gravada pela banda *The Rolling Stones* em 1965).

desvinculação do componente verbal (a letra das músicas) do componente acústico (a melodia). Muitos estudos focam um ou outro componente, mas não os dois conjuntamente (BERTOLI-DUTRA, 2010; 2014). 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). Por fim, a quarta tendência é a correlação entre a letra de música e a parte acústica da música (MAYER; NEUMAYER; RAUBER, 2008).

Há, portanto, quatro lacunas significativas na literatura prévia. Uma relativa à escassez de estudos empíricos, uma segunda que diz respeito à escassez de estudos envolvendo tanto o componente verbal quanto o acústico. A terceira é concernente à carência de estudos baseados em grandes amostras de dados e a última refere-se à falta de uma correlação musical que leve em conta as características linguísticas das letras de música e as características acústicas da produção musical, simultaneamente.

Tendo em vista essas lacunas, a presente tese 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 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 semânticas, ambas capazes de revelar padrões co-ocorrentes 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.

Desse modo, podemos dizer que o objetivo geral da presente tese é revelar as dimensões multimodais, multidimensionais, de base *big data*, da música popular em inglês. Os objetivos específicos são:

1. Identificar as características léxico-semânticas da música popular em inglês, por meio de dimensões de variação, com o emprego da Análise Multidimensional centrada na semântica da língua (BERBER SARDINHA, 2021);

2. Mensurar as características sonoras da música popular, por meio da compilação de métricas para cada uma das músicas do corpus, tais como dançabilidade, energia, tom, volume e modalidade. Para cada métrica, foi obtido um valor numérico,

gerado pela plataforma Spotify, que propiciou uma Análise Multidimensional que investiga o modo acústico.

3. Identificar a correlação entre as análises semântica e acústica por meio de uma correlação canônica, que tem como principal objetivo buscar a relação entre dois conjuntos de variáveis encontrando combinações lineares para cada um dos conjuntos, de modo a maximizar as correlações positivas entre os grupos.

Para atingir esses objetivos, o presente estudo responderá às seguintes perguntas de pesquisa:

1. Quais as dimensões semânticas do registro música popular em língua inglesa?
2. Quais as dimensões acústicas do do registro música popular em língua inglesa?
3. Quais as dimensões multimodais do registro música popular em língua inglesa?

A fim de levar a cabo os objetivos elencados, a presente tese emprega a 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, 2004; BIBER; CONRAD; REPPEN, 1998; MCENERY; WILSON, 1996). Tais amostras, chamadas de *corpora*², 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, 2020; SINCLAIR, 1991; SINCLAIR; JONES; DALEY, 1970/2004).

Com o aparato da Linguística de Corpus, nesta pesquisa foi coletado e processado um *corpus* de grandes proporções, de músicas em língua inglesa, bem como seus metadados, constituídos tanto das características acústicas das músicas quanto de informações como o ano de lançamento da música e os artistas envolvidos. A pesquisa aqui proposta é amparada pela missão do Programa de Linguística Aplicada e Estudos da Linguagem (LAEL) da PUC-SP, que compreende a promoção da inovação, da internacionalização e da ciência aberta. O LAEL concebe os Estudos da Linguagem como

² Plural de corpus.

campo de pesquisa nas suas múltiplas modalidades – escrita, falada, acústica, visual, háptica e sinalizada, bem como nos seus cruzamentos.

Mais especificamente, a pesquisa viabiliza a Dimensão 4 do Plano Estratégico do Programa Inovação e Ciência Aberta, consoante com o objetivo de implementar a ciência aberta e a inovação como princípios norteadores da política de pesquisa do Programa, assim como as metas de disponibilizar recursos de pesquisa, e de produzir e disponibilizar software para análises linguísticas. Portanto, a presente pesquisa inscreve-se em uma proposta inovadora tanto do ponto de vista científico quanto tecnológico, com potencial de impacto na pesquisa linguística e musical, que contribui para a internacionalização, avanço tecnológico e inovação da Universidade. Os produtos e dados foram disponibilizados no Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades (BERBER SARDINHA; MADUREIRA; BRAIT; SOUZA-E-SILVA et al., 2021), projeto institucional do PEPG em LAEL que tem o apoio do CNPq (Chamadas CNPq 25/2020 e 02/2021). Os insumos e materiais da pesquisa também foram disponibilizados na plataforma *Open Science Framework* (OSF). Ambas as iniciativas estão em consonância com o plano estratégico do PEPG em LAEL, que prevê a ciência aberta como política de pesquisa do Programa.

A presente tese está dividida em sete capítulos principais. A fundamentação teórica da tese vem na sequência da introdução e se subdivide em três partes: a linguística de corpus, focando principalmente na análise multidimensional e análise de correlação canônica do registro musical, os aspectos multimodais que aliam aspectos sonoros às letras de música, além de uma retrospectiva do cenário musical nos últimos 100 anos. A metodologia é o terceiro tópico, que analisará questões como o corpus da pesquisa, a análise multidimensional semântica das letras de música, a análise multidimensional da parte acústica das músicas e a análise de correlação modal do registro musical. A quarta parte apresenta os resultados e a discussão dos dados, seguida das considerações finais (quinta parte), das referências bibliográficas (sexta parte) e dos anexos (sétima parte).

2. FUNDAMENTAÇÃO TEÓRICA

*So when I'm all choked up and I can't find the words*³.

Lady Gaga

A música faz parte da nossa vida desde muito cedo. Todos temos uma relação intrínseca com a música, pois ela nos remete ao terreno das emoções e, frequentemente, traz consigo uma memória afetiva desse registro, seja pelo som de uma música, pela letra ou até mesmo por uma frase que fica grudada na nossa mente⁴. Para Lull (1992, p. 1), “a música é uma sequência extraordinária de pensamentos e de emoções, que expressa significados de modo único, sem paralelo na vida humana”. Dessa forma, um registro tão atrelado ao nosso dia a dia como a música pode produzir “sentidos” e transmitir significados por vezes pouco explorados.

A proposta da fundamentação teórica da presente pesquisa é mostrar como trabalhos anteriores lidaram com o registro música em língua inglesa, tanto em relação ao modo verbal (letra de música) quanto ao modo acústico.

A Linguística de Corpus forma a base teórica-metodológica do presente trabalho, com ênfase na análise multidimensional em suas vertentes gramatical/funcional e semântica, que envolve variação de campos semânticos, e os aspectos sonoros analisados nesta pesquisa através da análise multidimensional acústica. Na sequência, passaremos a explicitar as bases da correlação canônica multimodal.

Por fim, mas não menos importante, será examinada a trajetória histórica da música popular norte-americana, encerrando a fundamentação teórica.

³ “Então quando estou toda engasgada e não consigo encontrar as palavras” (trecho da música *Always remember us this way* gravada pela cantora Lady Gaga em 2018).

⁴ Parafrazeando Murphey (1990) com o fenômeno “the song stuck in my head”.

2.1 LINGUÍSTICA DE *CORPUS*

A Linguística de *Corpus* é a área da Linguística Aplicada que trabalha com a coleta e análise de corpus (plural *corpora*). Ela se ocupa de dados linguísticos produzidos por falantes reais, coletados para serem objeto de pesquisa (BERBER SARDINHA, 2004), ou seja, os textos falados ou escritos devem ser autênticos. Cabe ao pesquisador coletar esses dados do mundo real e utilizá-lo para a sua pesquisa. Neste trabalho, coletamos músicas em língua inglesa para utilizá-las como corpus de estudo.

Pode-se dizer que a Linguística de *Corpus* surgiu a partir da necessidade dos linguistas em utilizar dados reais para poderem entender o funcionamento linguístico. Porém, para um trabalho dessa proporção, necessitava-se de uma grande quantidade de textos que representasse uma língua ou uma variedade textual, o que seria inviável de ser feito manualmente sem incorrer em erros. De certa maneira, podemos atrelar o *boom* da Linguística de *Corpus* ao surgimento do computador, dado que apenas a partir dele uma grande quantidade de textos pode ser armazenada e analisada de forma confiável.

Atualmente, entre as ferramentas computacionais mais comuns à disposição dos linguistas de corpus e que foram (entre outras) utilizadas nesta tese, destacam-se três:

- (1) programas geradores de listas de palavras, que fazem a contagem das palavras em um *corpus* – o desta pesquisa, por exemplo, possui 103.809.650 palavras);
- (2) concordanciadores, que permitem que o usuário procure por palavras específicas em um corpus. Por exemplo, com base no *corpus* desta pesquisa, ao digitar a palavra “love” no concordanciador, dos inúmeros resultados encontrados, foram selecionados dois: *All you need is love* e *I was born to love you with every single beat of my heart*. O concordanciador fornece a palavra dentro do horizonte lexical em que ela se encontra e não apenas isoladamente. Nos exemplos aqui citados, vemos o uso do termo de busca (palavra nódulo) com dois sentidos diferentes: o primeiro, como algo necessário; e o segundo, como função única da vida do interlocutor. Há também duas funções sintáticas diferentes: no primeiro caso, a palavra é um substantivo; e no segundo, um verbo;
- (3) etiquetadores, que inserem etiquetas (códigos) de origem morfossintática, lexical ou semântica, como na presente pesquisa, em cada uma das palavras do corpus. Esse passo é extremamente necessário para pesquisas com análise de *big data* baseada em análise fatorial, como é o caso deste trabalho.

A Linguística de *Corpus* utiliza uma abordagem empirista cujo ponto central é a noção de que a língua é um sistema probabilístico, em que os traços linguísticos não ocorrem de forma aleatória e as palavras não se unem ao acaso (*language is never, ever, ever, random*⁵). Portanto, é possível evidenciar e quantificar padrões presentes em uma língua ou em uma variedade linguística. Há uma correlação entre os traços linguísticos e os contextos situacionais de uso da linguagem, ou seja, o “registro sempre importa”⁶. Podemos dizer que a Linguística de *Corpus* descortinou uma nova maneira de se estudar a linguagem e suas descobertas contribuem para a pesquisa em áreas muito diversas da Linguística, tais como Lexicografia, Ensino-aprendizagem de língua estrangeira e materna, Tradução e Linguística forense, entre outras.

Podemos citar algumas áreas já consolidadas dentro da Linguística de *Corpus* e outras que surgem a cada dia: (1) compilação de *corpora*; (2) desenvolvimento de ferramentas para análise de *corpora*; (3) descrição de linguagem; (4) uso de descrições baseadas em *corpora* para aplicações em ensino-aprendizagem de línguas, processamento de linguagem natural por máquinas, reconhecimento de voz e tradução.

Biber, Conrad e Reppen (1998) afirmam que uma das características da abordagem com base em *corpus* é que ela se baseia em técnicas de análise quantitativas e qualitativas. As interpretações dos dados quantitativos são essenciais para explicar as razões pelas quais existem certos padrões de ocorrência de uma palavra, sem se valer da intuição. Leech (1992) afirma que não há como utilizar a intuição do falante nativo em testes probabilísticos. Desse modo, com os resultados das interpretações quantitativas, devemos examinar os textos mais representativos do corpus (pois a análise quantitativa nos fornece este dado) e, em seguida, realizar uma análise qualitativa.

Essa união das perspectivas quantitativa e qualitativa é muito bem explorada na Análise Multidimensional, com o objetivo de caracterizar uma língua, ou um conjunto de tipos de textos, em articulação com características situacionais dos textos, que revelam onde, por quem, para quem e por que aquele texto foi produzido. Essa abordagem será examinada em maior detalhe na próxima seção.

⁵ “A língua nunca, nunca, é aleatória” (Adam Kilgarriff, 2005).

⁶ “Register always matters” (Douglas Biber, 2013).

2.1.1 Análise Multidimensional

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. Esses conjuntos são identificados por meio de uma análise fatorial – cujos fatores resultantes, ao serem interpretados comunicativamente, passam a ser chamados de dimensões.

Antes da Análise Multidimensional, era comum o estudo linguístico centrado na co-ocorrência de poucos traços e a interpretação era basicamente intuitiva. A variação entre os registros⁷ era investigada com poucos parâmetros, geralmente de maneira dicotômica (texto “formal” ou “informal”) e incompleta, por privilegiar apenas uma das muitas diferenças que podem existir entre os textos.

A Análise Multidimensional teve seu início na década de 1980, nos Estados Unidos, com o linguista Douglas Biber. Desenvolvida no que se convencionou chamar de escola americana da Linguística de *Corpus*, a Análise Multidimensional vem a ser uma nova maneira de enxergar a língua, debruçando-se sobre vários textos ao mesmo tempo, e valendo-se de uma abordagem estatística, que envolve múltiplas variáveis – daí o nome Análise **Multidimensional** (AMD).

Com a AMD podemos trabalhar uma quantidade maior de parâmetros, que permitem uma comparação mais abrangente e com o maior número possível de características linguísticas. Como dito, o estudo se baseia em *corpora*, ou seja, um grande número de textos autênticos, tem um processamento essencialmente computacional, utilizando ferramentas para etiquetagem das características pesquisadas, e compara textos e registros de maneira a permitir uma descrição multifacetada e fundamentada em dados.

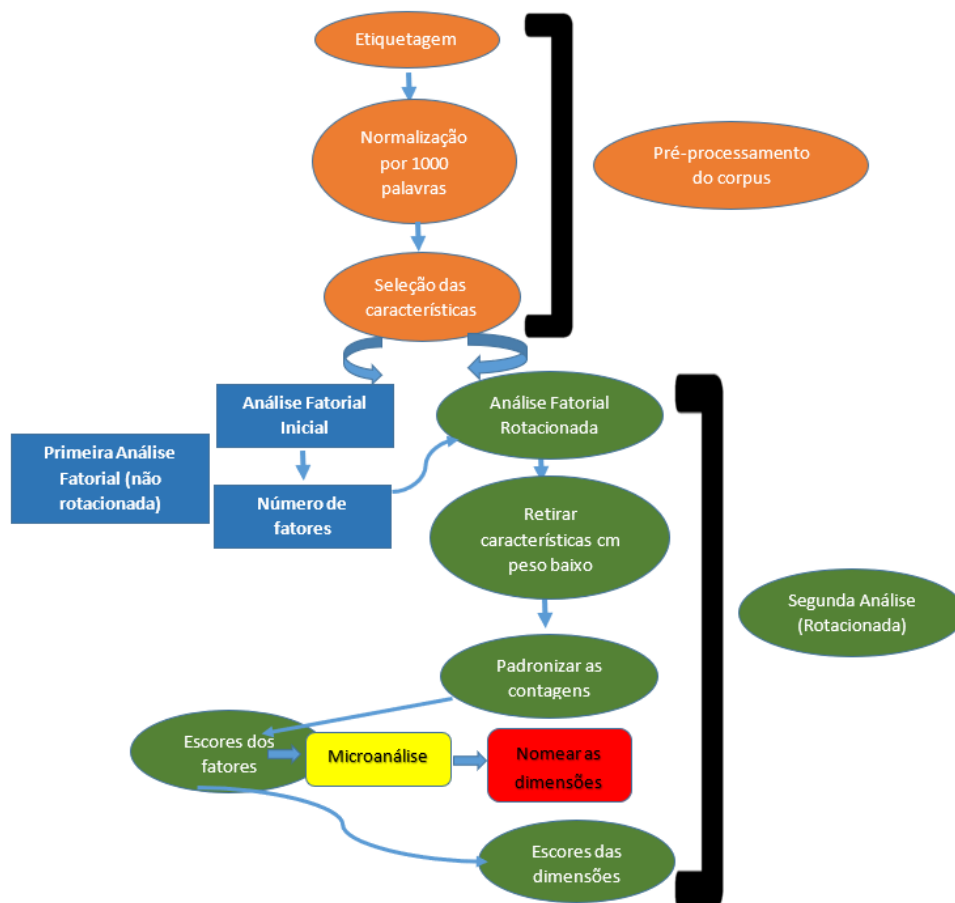
Antes do advento da AMD, a Linguística de *Corpus* valia-se dos padrões presentes nas colocações, coligações e preferência semântica. A AMD vem a ser um passo além, na medida em que se vale da variação presente nesses padrões, mas numa quantidade de textos muito maior e, mostrando padrões de co-ocorrência em diferentes registros. Assim, o que é comum em um registro como “horóscopo” pode estar presente em um “artigo

⁷ Registro é uma “variedade linguística definida por aspectos situacionais, incluindo o propósito do falante, a relação entre falante e ouvinte, e o contexto de produção.” (BIBER, 2009, p. 823).

científico”, pois em ambos os registros ocorrem características que compartilham as mesmas funções da língua (BERBER SARDINHA; KAUFFMANN; ACUNZO, 2014).

Apesar de fundamentalmente a AMD ser uma abordagem quantitativa, ela não prescinde da análise qualitativa, na medida em que, para nomear as dimensões, é necessário efetuar uma interpretação qualitativa dos fatores. Conforme mencionado, a AMD se fundamenta na análise fatorial cujo resultado são fatores que, ao serem interpretados no textos do corpus, recebem o nome de dimensão. A Figura 1, que ilustra os passos para a realização de uma AMD, possui três partes fundamentais: (1) o pré-processamento do corpus, incluindo a etiquetagem, a normalização e a seleção das variáveis (traços linguísticos) a serem estudadas; (2) uma primeira análise fatorial (não rotacionada), em que o analista decide o número de fatores a serem extraídos através da análise do gráfico de escarpa; (3) e a segunda análise fatorial (rotacionada), que envolve extrair o número final de fatores, retirar variáveis com peso baixo nos fatores, realizar o cálculo de escores e, por fim, nomear as dimensões, através de microanálises.

FIGURA 1 – PASSOS DE UMA ANÁLISE MULTIDIMENSIONAL



Fonte: Delfino (2021)

Com o advento da Análise Multidimensional, passou a ser possível descrever registros diferentes sob uma base comum, destacando funções comunicativas que os une ou que os separa. Biber apresentou as dimensões funcionais/gramaticais da língua inglesa e os trabalhos que se seguiram (BIBER, 1988; BIBER; KURHIAN, 2007, entre outros) focaram nessa perspectiva.

Alguns trabalhos de Análise Multidimensional utilizaram a música pop como objeto de estudo. Werner (2021) usou um corpus de canções pop para fazer um mapeamento do registro música pop, tendo as dimensões originais de Biber (1988) como referência. Por sua vez, Delfino (2016), com uma amostra diferente da coletada por Werner, também fez um mapeamento de um corpus de músicas norte-americanas e britânicas para alocar as letras de música nas dimensões da língua inglesa segundo Biber (1988). Já Bértoli-Dutra (2014) fez uma análise funcional/gramatical com um corpus de músicas norte-americanas e britânicas entre 1940 e 2009, com o objetivo de investigar a variação na música pop escrita em língua inglesa, através da extração das dimensões da música pop no período estudado. Apesar de também terem utilizado a AMD como metodologia comum, esses trabalhos precedentes não levaram em consideração as variáveis tratadas na presente pesquisa, que enfocam a parte lexical do registro letra de música, bem como as características acústicas de tais músicas.

Estudos com técnicas de estatística multivariada para a aquisição de conhecimento lexical avançaram na década de 1990. Bindi et al. (1991), através de uma análise chamada Dimensionamento Multidimensional (MDS), estudou as relações entre tipos de palavras semanticamente relacionadas (por exemplo, *piccolo* e *corto*). Em 1993, Biber usou a análise fatorial para investigar a padronização no uso de colocações e percebeu que “each factor represents a separate grouping of collocations that tend to co-occur frequently in texts”⁸ (BIBER, 1993, p. 536).

A seguir, outros estudos expandiram o escopo original da AMD. Salsbury et al. (2011), por meio de testes estatísticos ANOVA (análise de variação), perceberam que o léxico de aprendizes de inglês como segunda língua torna-se mais refinado e abstrato com

⁸ “cada fator representa um grupo único de colocações que tendem a coocorrer com frequência nos textos” (tradução nossa).

o passar do tempo (o intervalo temporal era a variável independente e os valores contidos na base de dados do conselho de pesquisa médica, as variáveis dependentes). A partir de 2014, Berber Sardinha propôs estudos com léxico utilizando a AMD para identificar padrões de variação lexical. O autor propôs três novos tipos de AMD: a AMD lexical/temática, a AMD colocacional e a AMD semântica.

O estudo de características lexicais na AMD já foi utilizado para analisar as dimensões do discurso religioso (VEIGA, 2021), a linguagem das mídias sociais (MILANEZ et al., 2022), as representações culturais (BERBER SARDINHA, 2020), a linguagem verbal das artes visuais (ROMEIRO, 2020), o estilo de escrita de Machado de Assis (KAUFFMANN, 2020), os temas subjacentes do periódico Revista Brasileira de Linguística Aplicada (DELFINO et al., 2018) e o histórico da Linguística Aplicada no mundo, através de vários periódicos da área (BERBER SARDINHA, 2017a). Estudos onde os traços linguísticos utilizados como variáveis de uma AMD são colocações⁹ foram realizados no intuito de revelar as dimensões colocacionais presentes no inglês americano (BERBER SARDINHA, 2017b) e no inglês acadêmico (ZUPPARDI, 2020).

Diversas perspectivas que expandem a metodologia da AMD foram apresentadas acima, sejam de natureza gramatical/funcional, lexical/temática, colocacional ou com a proposta da perspectiva semântica, como na presente tese. Esses estudos valem-se da união das análises quantitativa e qualitativa proporcionada pela AMD, levando o estudo linguístico para além do texto, para descrever um registro, uma variedade linguística ou mesmo uma língua.

Na medida em que é, até onde esta pesquisadora apurou, o primeiro estudo multidimensional onde os traços linguísticos são os dados de categorias semânticas, na presente tese discutiremos a seguir como esses campos semânticos são escolhidos, a partir de um etiquetador desenvolvido para capturar tais traços nos textos.

2.1.2 Anotação semântica

Estudos multidimensionais, sejam eles gramático-funcionais, lexical-temáticos, colocacionais ou semânticos, baseiam-se em análise fatorial e seguem, basicamente, os mesmos passos metodológicos. A diferença fundamental entre eles é a anotação utilizada,

⁹ Colocação vem a ser a co-ocorrência de duas (ou mais) palavras numa frequência maior do que seria de se esperar caso a co-ocorrência fosse aleatória.

ou seja, quais características presentes no *corpus* foram priorizadas e transformadas em variáveis para serem analisadas. Por exemplo, se o pesquisador pretende desenvolver uma análise multidimensional gramatical funcional, suas variáveis serão gramaticais, como pronomes, verbos e substantivos. Nesse caso, o corpus deve ser etiquetado com um etiquetador léxico-gramatical, como é o caso do Biber Tagger. Já se a análise for lexical-temática, cada palavra será etiquetada a partir do seu lema, e etiquetadores como o TreeTagger são indicados para tal passo. Na presente pesquisa, trabalhamos com dimensões semânticas, então necessitamos etiquetar o corpus com um etiquetador semântico, o qual é detalhado a seguir.

O modelo de anotação semântico utilizado nesta tese é o do Sistema de Análise Semântica UCREL (no original, UCREL *Semantic Analysis System*, sigla USAS), desenvolvido na Universidade de Lancaster¹⁰. Ele se baseia na classificação do léxico em 21 categorias maiores, denominadas de campos semânticos, representadas pelas letras do alfabeto, conforme mostra a Figura 2. Estes campos são adequados para trabalhar com vários tipos de textos. Archer et al. (2004) mostra as semelhanças nas taxonomias utilizadas pelo USAS e o dicionário Collins (2001). Na verdade, o etiquetador USAS trabalha com campos semânticos, ou domínios semânticos, que agrupam os sentidos das palavras relacionadas com o mesmo conceito. Por exemplo, o campo semântico *Education*¹¹ abarca palavras como *class, school, course, board, college, university, private, high (school), middle (school), classroom*¹², entre outras.

¹⁰ <http://ucrel-api.lancaster.ac.uk/usas/tagger.html>

¹¹ Educação.

¹² aula, escola, curso, lousa, faculdade, universidade, privada, ensino médio, ensino fundamental, sala de aula.

FIGURA 2 - CAMPOS SEMÂNTICOS DO TAGSET DO ETIQUETADOR USAS¹³

A General and abstract terms	B The body and the individual	C Arts and Crafts	E Emotional actions, states and processes
F Food and farming	G Government and the public domain	H Architecture, buildings, houses and the home	I Money and commerce
K Entertainment, sports and games	L Life and living things	M Movement, location, travel and transport	N Numbers and measurement
O Substances, materials, objects and equipment	P Education	Q Linguistic actions, states and processes	S Social actions, states and processes
T Time	W The world and our environment	X Psychological actions, states and processes	Y Science and technology
	Z Names and grammatical words		

Fonte: Archer et al. (2002, p. 2)

Os campos semânticos incluem uma categoria Z, que relaciona nomes próprios e palavras gramaticais. Essa categoria, assim como as outras, recebe números após a letra. No caso da categoria Z, ela se subdivide de Z0 (nome próprio não encontrado) a Z99 (palavra não relacionada com nenhuma etiqueta). No *corpus* desta pesquisa, por conter textos do inglês não padrão, com muitas gírias (por exemplo, *nigga*), a etiqueta Z99 foi muito presente e uma inspeção manual foi necessária¹⁴. Na Figura 3, ilustramos as categorias do campo semântico *Time* no USAS, com exemplos do *corpus* da pesquisa. A subcategoria T1.3, por exemplo, inclui substantivos (*night*, *weekends*), verbos (*ticking*), adjetivos (*short*) e preposições (*for*).

¹³ Tradução: A - Termos gerais e abstratos; B- O corpo e o indivíduo; C- Artes e artesanato; E - Ações, estados e processos emocionais; F - Comida e agricultura; G- Governo e domínio público; H - Arquitetura, prédios, casas e lar; I - Dinheiro e comércio; K - Entretenimento, esportes e jogos; L - Vida e coisas vivas; M - Movimento, localização, viagem e transporte; N - Números e medidas; O - Substâncias, materiais, objetos e equipamento; P - Educação; Q - Ações, estados e processos linguísticos; S - Ações, estados e processos sociais; T - Tempo; W - O mundo e nosso meio ambiente; X - Ações, estados e processos psicológicos; Y - Ciência e tecnologia; Z - Nomes próprios e palavras gramaticais.

¹⁴ Essa inspeção será detalhada na metodologia.

FIGURA 3 - CATEGORIAS E SUBCATEGORIAS DO CAMPO SEMÂNTICO *TIME* NO USAS

T1 Time
T1.1 General
T1.1.1 Past: <i>so long ago</i>
T1.1.2 Present: <i>now, today; yet</i>
T1.1.3 Future: <i>will; tomorrow; going</i>
T1.2 Momentary: <i>dawning; moment; five</i>
T1.3 Period: <i>night; ticking; weekends; short; for</i>
T2 Beginning/ending: <i>begin; still; started off</i>
T3 Old/new/young; age: <i>latest; fresh; grow</i>
T4 Early/late: <i>late; early</i>

Fonte: elaborada pela autora

Um exemplo de texto etiquetado com o USAS pode ser visto na Figura 4. A música *Save a Prayer*, da banda Duran Duran, de 1982 (arquivo 166414), está ilustrada com algumas características que o etiquetador nos traz. Assim, quando é processada no etiquetador USAS a frase *You saw me standing by the wall corner of a main street*, o etiquetador vai trazer como resultado cada palavra itemizada, para aparecer isolada em uma linha, apontando antes de cada palavra um código corresponde à categoria morfossintática e, após a palavra, um outro código que equivale ao campo semântico. Por exemplo, a segunda palavra do texto é “*saw*”, que possui antes dela o código VVD, que corresponde a um verbo no passado; na sequência, há vários códigos, pois o etiquetador coloca todas as possibilidades de campos semânticos que a palavra possa ter. Optamos, nesta pesquisa, por permanecer sempre com o primeiro código fornecido pelo etiquetador, por ser o que possui maior chance de acerto. No caso em questão, da palavra “*saw*”, utilizamos a etiqueta X3.4, que significa *Sensory: Sight* (sensorial: visão). Ou seja, o etiquetador fornece a categoria morfossintática da palavra e o campo semântico ao qual ela está inserida.

FIGURA 4 - TEXTO ETIQUETADO COM O USAS

PPY	You	Z8mf
VVD	saw	X3.4 X2.1 S1.1.1 X2.5+ X2.3+ X3 A7+ S3.2
PPIO1	me	Z8mf
VVG	standing	A1.1.1[i1.2.1 S8+[i1.2.1 M8[i1.2.1 M6 M2
II	by	A1.1.1[i1.2.2 S8+[i1.2.2 M8[i1.2.2 Z5
AT	the	Z5
NN1	wall	H2 O2 N5+ S8-
NN1	corner	O4.4 M6 K5.1
IO	of	Z5
AT1	a	Z5
JJ	main	A11.1+ N5+++
NN1	street	M3/H3
CC	And	Z5
AT	the	Z5
NN2	lights	W2 O3/W2 M3 X4.1

Fonte: elaborada pela autora

O processo de anotar o *corpus* com as etiquetas do USAS ocorre automaticamente, a partir de uma complexa combinação de léxico e regras, que incluem uma tokenização prévia e etiquetagem morfossintática e semântica (RAYSON, 2007) a partir do *Corpus Internacional de Inglês (International Corpus of English)* usando o USAS.

2.1.3 Multimodalidade: o acústico e o verbal na música

A música popular tal qual levada em conta em nosso estudo possui dois modos constituintes, o verbal (registrado na letra da música) e o acústico (registrado pelo acompanhamento dos instrumentos musicais). Os estudos até a presente data geralmente costumam levar em consideração um ou outro modo apenas. Podemos citar Barbosa (2012), Bértoli Dutra (2010; 2014); Delfino (2016) e Werner (2012) como trabalhos envolvendo apenas a letra da música, e Pinarbasi (2019), Hongpanarak e Mongkolnavin (2021), Pavlik (2019) e Panda et al. (2021) como trabalhos envolvendo apenas o modo acústico da música. Esta tese sublinha a indissociação desses dois modos na medida em que faz um estudo multidimensional semântico das letras concomitantemente a um estudo multidimensional de aspectos acústicos das melodias.

Na consecução dessa meta, operacionalizamos o estudo do modo acústico a partir de onze características acústicas calculadas computacionalmente pela empresa Spotify para as canções comercializadas. O serviço por assinatura Spotify, plataforma de streaming de música utilizada nesta pesquisa, foi lançada em 2008 por uma *startup* de Estocolmo, na Suécia. A ideia era se tornar a trilha sonora de seus assinantes, oferecendo

uma grande variedade de playlists e podcasts todas as manhãs. Em janeiro de 2022, foi considerado o maior serviço de streaming de música do mundo, com 381 milhões de usuários e mais de 70 milhões de canções¹⁵. A plataforma também sugere músicas que o ouvinte pode gostar, a partir de algoritmos que trabalham com o histórico de dados de consumo musical do assinante, que ela armazena.

Além de analisar os dados de usuários, a plataforma também realiza uma análise das músicas que estão armazenadas no serviço por meio de algumas métricas que demonstram em que medida uma música pode ser considerada um hit. Ochi et al. (2021) analisaram a conexão entre a popularidade das músicas e a dançabilidade, que vem a ser uma das métricas que o Spotify utiliza para mapear as músicas incluídas na plataforma, através de uma combinação de elementos musicais incluindo ritmo, estabilidade, força da batida e regularidade como um todo. Esses valores variam de 0 a 1 e músicas mais próximas de 1 são consideradas mais dançantes.

Uma outra característica analisada pelo Spotify é a energia da música, isto é, uma medida perceptiva da intensidade e atividade baseada em uma escala dinâmica que envolve a altura da música, o timbre, o início de uma nota musical e a entropia de maneira geral. Em outras palavras, é o senso de movimento na música, o que quer que cativa o ouvinte. Em músicas altas, a energia musical é mais fácil de ser identificada. Quanto mais a bateria é utilizada e mais alto o artista canta, mais alta é a energia da música. O Spotify calcula essa métrica através da amplitude dinâmica (a diferença entre o ponto mais alto e mais baixo da música), o volume captado e o timbre da música. Esses valores também variam de 0 a 1 e músicas mais próximas de 1 são consideradas extremamente energéticas.

Já a valência descreve a positividade musical transmitida. Músicas com valência alta soam mais positivas, eufóricas. Os valores também variam de 0 a 1, então quanto mais próximo de 1 mais positiva é a música.

Como o nome sugere, a acústica mostra se a faixa é ou não acústica, isto é, se a faixa é mais instrumental, priorizando a parte melódica, com pouca ou nenhuma letra. Os valores também variam de 0 a 1 e quanto mais próxima a música está de 1, maior chance de ela ser uma música instrumental. Panda et al. (2021) fizeram experimentos com as métricas do API (*Applications Protocol Interface*¹⁶) do Spotify para classificar as músicas

¹⁵ <https://newsroom.spotify.com/company-info/>

¹⁶ Interface de programação de aplicações.

por meio da emoção e chegaram ao resultado de que a energia, a valência e a acústica são altamente relevantes. O fato de a acústica estar relacionada com a energia é algo novo.

A característica *key* (tom) do Spotify mostra a qualidade do som através de vibrações produzidas pelo volume alto ou baixo de um tom – melhor explicando, mostra a escala maior ou menor em que a música opera. Provavelmente uma música com escala maior será mais dançante e mais feliz, enquanto uma música com escala menor tende a ser mais triste.

A medida volume do Spotify refere-se aos valores em decibéis da faixa e variam entre -60 e 0dB. Pinarbasi (2019) fez um estudo com as mesmas métricas do Spotify usadas na presente pesquisa para estudar o mercado fonográfico da Turquia e chegou a resultados que mostram que as músicas mais populares naquele país geralmente possuem mais energia, volume e dançabilidade. O mesmo estudo demonstrou correlações positivas entre energia e volume, correlações medianas entre valência, dançabilidade, energia e volume, mas correlações negativas entre energia e musicalidade, volume e musicalidade. O referido autor chegou à conclusão, através de um teste estatístico chamado árvore de decisão, que coloca em um ranking as variáveis do Spotify mais importantes para a classificação das músicas por popularidade. Em tal categorização, as variáveis que mais se destacaram foram energia, volume, musicalidade, tempo e dançabilidade, sendo volume a característica mais significativa para classificar as músicas do *corpus* (composto por 676 músicas). Para o estudo, a combinação ideal para uma música ter chance de vir a ser um hit de sucesso deve conter volume de -6,5dB, energia com valores menores que 0,64 e a duração da música menor que 224 segundos (menos que 4 minutos).

Outra variável do Spotify usada na presente pesquisa foi a modalidade de uma faixa, que pode ser maior ou menor, com apenas duas medidas, 0 e 1. A literatura apresenta trabalhos como o de Panda et al. (2021) relatando que valores altos de modalidade tendem a estar associados com felicidade. Intuitivamente, espera-se que uma métrica como dançabilidade esteja correlacionada com modalidade, o que não ocorreu, talvez pelo fato de o algoritmo usado pelo Spotify não ser robusto, ou por essa variável trabalhar melhor em consonância com outras características não estudadas. A correlação entre dançabilidade e modalidade também não ocorreu na presente pesquisa, na medida em que as músicas com alta modalidade tiveram a tendência de serem músicas mais alegres, em tom alto e com sonoridade menos emotiva, mas não dançante, visto que dançabilidade está altamente relacionada à valência e não modalidade (vide Seção 4.2.3).

A medida do Spotify denominada verbosidade detecta a presença de palavras faladas (e não cantadas) em uma faixa. Quanto mais os valores estão próximos de 1, maior o valor da verbosidade, e mais a faixa é parecida com um discurso – por exemplo, como ocorre com talk shows, audiobooks e na poesia). Músicas de rap tendem a ser bons exemplos de faixas com alta verbosidade. Hongpanarak e Mongkolnavin (2021) estudaram as métricas do Spotify para relacionar oito das métricas estudadas na presente pesquisa com a personalidade de voluntários que concordaram em compartilhar o que ouviam no Spotify durante um período de três meses e chegaram a conclusão de que tanto a energia como a verbosidade se correlacionam negativamente com a extroversão e positivamente com a acústica. Ou seja, participantes mais extrovertidos tendem a ouvir mais músicas acústicas e a ouvir bem menos faixas rápidas, altas ou barulhentas. O mesmo estudo detectou que aqueles que são mais abertos a experiências novas possuem correlação positiva com vivacidade, que vem a ser uma medida que detecta a presença de público na gravação. Valores mais altos de vivacidade indicam uma probabilidade de que a faixa seja uma performance ao vivo. A partir de tal estudo, pode-se concluir que pessoas abertas a novas experiências tendem a ouvir música ao vivo.

Hongpanarak e Mongkolnavin (2021) também concluíram que pessoas mais amáveis possuem uma correlação negativa com o ritmo, que vem a ser a velocidade de uma faixa, derivada da média da duração da batida. Medida essa feita em BPM (batidas por minuto). Em outras palavras, os autores concluíram que pessoas mais amáveis tendem a gostar de músicas com batidas lentas, tais como baladas. Já a instrumentalidade se relaciona com a quantidade de vocais que a faixa possui (“lalala” é considerado instrumental). Músicas que possuem mais do que 0,5 (numa escala que vai de 0 a 1) de valor de instrumentalidade tendem a ser faixas instrumentais.

Pavlik (2019) estudou essas medidas acústicas do Spotify através de procedimentos estatísticos como árvore de decisão, floresta de decisão aleatória e XG Boost (um algoritmo usado em *machine learning*) e, com uma acurácia de 54%, conseguiu relacionar algumas medidas acústicas com gêneros musicais. Segundo o estudo, o rap é ligado à verbosidade. Já o rock e a música latina se opõem em relação à variável dançabilidade – ou seja, músicas latinas são muito dançantes, enquanto canções de rock não são para dançar. A música eletrônica possui uma quantidade alta de batidas por minuto, e o R&B está ligado a músicas maiores (não analisamos o tamanho da música na presente pesquisa). A música pop não se enquadrava em nenhuma das medidas

acústicas, o que é curioso porque se as classificações que existem baseiam-se no componente acústico. Desse modo, uma música é considerada pop por outros atributos que não os de natureza acústica, de acordo com a autora.

Com as dimensões semânticas e acústicas reveladas, podemos detalhar como se dá através da análise canônica a correlação entre esses pares de dimensões, capazes de nos levar a grupos multimodais que atuam no registro musical.

2.2 ANÁLISE DE CORRELAÇÃO CANÔNICA

A análise de correlação canônica proporciona o estudo da relação entre dois conjuntos de variáveis (TABACHNICK; FIDELL, 2007). De acordo com as autoras, “a correlação canônica proporciona uma análise estatística para a pesquisa, na qual cada item é medido em dois conjuntos de variáveis e o pesquisador deseja saber se e como os dois conjuntos relacionam-se entre si.”¹⁷ (TABACHNICK; FIDELL, 2007, p. 571).

Nesta pesquisa, nosso objetivo foi fazer uma categorização das músicas em inglês através das dimensões semânticas das letras das músicas e das dimensões acústicas, observando a relação existente entre elas. Mayer (2018) empregou essa metodologia para fazer a correlação entre as dimensões funcionais/gramaticais e lexicais/temáticas de um corpus de comentários da internet para saber o que e como se escreve na *web*. Kauffmann (2020) também se valeu da correlação das dimensões funcionais/gramaticais e lexicais/temáticas para descobrir o estilo presente nos textos de Machado de Assis. Berber Sardinha (2021) inovou ao utilizar a correlação canônica para fazer um estudo multimodal, na medida em que relacionou dimensões de textos e dimensões de imagens de um corpus de tweets. Seguindo os passos de Berber Sardinha (2021), realizamos um estudo multimodal das músicas relacionando, através da análise canônica, dimensões semânticas das letras de música e dimensões de elementos acústicos presentes nas músicas.

Em uma análise canônica, cada conjunto de variáveis é combinado para produzir um valor de predição que tenha uma correlação alta com o outro conjunto (TABACHNICK; FIDELL, 2007). Assim, a princípio, são determinadas combinações

¹⁷ No original: “canonical correlation provides a statistical analysis for research in which each subject is measured on two sets of variables and the researcher wants to know if and how the two sets relate to each other.” (TABACHNICK; FIDELL, 2007, p. 571).

lineares de cada conjunto separadamente, que serão aqui chamadas de variáveis canônicas. Em seguida são determinados pares de combinações canônicas, que chamaremos de correlações canônicas.

Segundo Tabachnick e Fidell (2007), os benefícios da correlação canônica são o fato de lidar com a dimensionalidade e permitir uma estrutura para a análise de variáveis múltiplas em cada lado de uma equação linear. Para realizar essa correlação, Mayer (2018) afirma que os dados devem estar normalizados, ter uma relação linear e as relações entre os pares de variáveis devem ser homocedásticos, ou seja, com igual dispersão ou variância (os dados desta pesquisa se encaixam em tais requisitos).

A análise canônica consiste nos seguintes passos (TABACHNICK; FIDELL, 2007, p. 578-581):

- (1) Geração de uma matriz de correlação em quatro partes: entre o primeiro grupo de variáveis, entre o segundo grupo de variáveis e as duas matrizes de correlações entre o primeiro e o segundo conjunto de variáveis.
- (2) Cálculo dos *eigenvalues* e *eigenvectors*, para a redistribuição em pares de variáveis canônicas, de forma que cada par capture uma grande variância compartilhada por combinações lineares dos conjuntos de variáveis. A solução para a obtenção dos *eigenvalues* redistribui a variância na matriz, de maneira a formar as variáveis canônicas, em oposição a diversas variáveis individuais. O *eigenvector* que corresponde a cada *eigenvalue* é transformado em coeficientes usados na combinação das variáveis para a combinação das variáveis canônicas.
- (3) Cálculo dos coeficientes canônicos de um dos conjuntos, que são produtos da matriz de correlação e a matriz dos *eigenvectors*.
- (4) Cálculo dos coeficientes do segundo conjunto, produto da matriz de correlação do segundo conjunto, matriz de correlação entre os dois conjuntos e a matriz formada por coeficientes para o primeiro conjunto.

A análise de correlação canônica permite, em suma, considerar ambos os conjuntos de dimensões resultantes das análises fatoriais efetuadas na AMD, gerando uma interpretação complexa que se superpõe a uma simples enunciação de cada análise,

isoladamente – como poderá ser visto na Seção xxx. A seguir, será traçado um panorama histórico da música pop em inglês, objeto de estudo desta pesquisa.

2.3 CONSIDERAÇÕES SOBRE A MÚSICA POPULAR EM LÍNGUA INGLESA

A música é um produto cultural e, como tal, influencia e é influenciada pela sociedade e por movimentos históricos que acompanham o momento em que é criada. O presente trabalho possui como corpus músicas em língua inglesa, de 1913 a 2018, perfazendo quase todo o século XX, passando pelas duas guerras mundiais, o nascimento do rock-and-roll concomitante ao surgimento do cinema falado com trilha sonora contendo músicas intrinsecamente ligadas ao roteiro de vários filmes desde então. O *corpus* adentra nos anos 1960 com movimentos culturais, tais como o *flower power* e os festivais atrelados a esse movimento, a era disco nos anos 1970, os anos 1980 com o estabelecimento da música pop e a entrada no século XXI.

A fim de relacionar a história da música popular com os acontecimentos histórico-culturais que ocorreram no mundo e influenciaram a música, seguem-se alguns fatos importantes para o cenário musical e o pano de fundo sociohistórico do mundo de então.

Começaremos no início do século XX, com a primeira guerra mundial, que inspirou tanto músicas patrióticas de apoio à guerra (*Keep the Home Fires Burning (Till The Boys Come Home)*, de 1914) como canções ativistas antiguerra (*I Didn't Raise my Boy to be a Soldier*, de 1915), expressando debates, ansiedades (*Lay down your arms*, de 1915), identidades sociais e mudanças no papel do homem e da mulher na sociedade (MULLEN, 2020).

Originado no final do século XIX como uma fusão de influências anglo-americana, africana e crioula no caldeirão cultural de Nova Orleans na Louisiana, o jazz teve a sua década de ouro nos anos 1920, com Louis Armstrong e Duke Ellington entre os músicos mais famosos. O solo e swing da música *Hotter than that* (1927), de Armstrong, personifica o que o jazz representou no período (GABBARD, 1995).

Durante as décadas de 1930 e 1940, a música country tornou-se muito popular devido aos filmes de faroeste veiculados por Hollywood (AUSTIN, 1984), que tinham

cantores como Roy Rogers e Gene Autry. Músicas como *The last round up* (1934), gravada por Gene Autry, fizeram tanto sucesso que, em 1944, a revista *Billboard* trocou o termo *folk songs and blues* por *country*.

Na década de 1940, o *blues* tornou-se eletrificado e transformou-se em *rhythm and blues* (R&B), já formando a base para o surgimento do rock-and-roll nas décadas seguintes (RIPANI, 2006). Um dos cantores mais representativos dessa era foi Louis Jordan, com músicas como *What's the use of getting sober (when you gonna get drunk again)* (1942).

O rock-and-roll dominou a música popular nos Estados Unidos e se espalhou pelo mundo a partir da metade da década de 1950 (ALTSCHULER, 2003). Sua origem recaí numa mistura de vários gêneros da black music, como o *rhythm and blues* (R&B), música gospel, country e pop. Os anos 1950 viram a alta expansão da guitarra elétrica, o que contribuiu para a chegada do rock, principalmente devido à Chuck Berry, que introduziu solos de guitarra em músicas como *Maybellene* (1955). Considerada a pioneira do rock-and-roll, com um tempo de 140 batidas por minuto, a canção era muito maior do que a maioria das músicas compostas anteriormente e promoveu mudanças no ritmo de duas batidas para quatro, “o que faz você querer dançar ou dirigir mais rápido”¹⁸ (CAMPBELL, 2018). Artistas como Elvis Presley, Chuck Berry, Bill Halley, Little Richard e Jerry Lee Lewis, entre outros, lançaram os primeiros hits do que seria o rock-and-roll. *Rock around the clock* (1955), de Bill Halley and His Comets, inauguraram a era rock (NILE, 1992) – essa música abria o filme *The Blackboard Jungle*, o que causou sensação nas plateias.

De acordo com Grossberg (1992, p. 172), o baby boom foi

decisivo para a rearticulação das estruturas da cultura popular depois da Segunda Guerra Mundial e da Guerra da Coreia; entre 1946 e 1964, nasceram 77 milhões de crianças e, em 1964, 40% da população dos Estados Unidos tinha menos de vinte anos.

O baby boom e o nascimento de um mercado jovem transformaram a juventude no público-alvo da indústria cultural:

¹⁸ A música fala sobre traição e carros de corrida.

os adolescentes norte-americanos do pós-guerra desfrutaram de uma abundância sem precedentes. O gosto cinematográfico, musical, literário e de entretenimento foi impulsionado pelo seu enorme poder aquisitivo, que as gravadoras e os produtores cinematográficos satisfaziam rapidamente. (WELSH, 1990, p.3).

Nos anos de 1950, as configurações específicas das circunstâncias forjaram uma aliança entre a “juventude” (grupo demográfico constituído por uma maioria de pessoas jovens) e o rock, transformando-os em sinônimos (ALTSCHULER, 2003). O rock-and-roll foi um estilo que atraiu tanto os jovens negros como os brancos e, culturalmente falando, influenciou movimentos pelos direitos civis, além da moda não apenas nos Estados Unidos, mas também a classe trabalhadora na Inglaterra, em Londres e em cidades mais periféricas e portuárias como Liverpool (GRANDE, 2006).

Diferentemente do que aconteceu na década de 1950, que foi basicamente dominada pelo rock-and-roll, os anos 1960 viram o surgimento de vários movimentos musicais, alavancando principalmente o que viria a ser conhecido mais tarde como *pop music*. A música *People Get Ready* (1965) foi considerada o hino dos direitos civis e regravaada por vários artistas retratando a mudança que a década de 1960 representou. O período trouxe transformações não apenas ao cenário musical, mas à cultura e à sociedade americana como um todo, que depois acabou se espalhando pelo mundo (EYERMAN; JAMISON, 1995).

Ao mesmo tempo, na década de 1960, ocorreu a “invasão britânica”, que vem a ser um termo usado principalmente nos Estados Unidos para descrever o grande número de bandas e cantores de rock-and-roll e pop britânicos que se tornaram populares nos Estados Unidos, principalmente entre 1964 e 1966 (OLWIG, 2010). Esse termo (invasão britânica) transformou-se em Beatlemania, depois que o grupo The Beatles se apresentou no programa de TV *The Ed Sullivan Show*, cantando a música *Twist and Shout* (1963). De acordo com Olwig (2010), entre 1964 e 1966 outros artistas (The Animals, Petula Clark, Herman’s Hermits, The Rolling Stones, The Who, entre outros) também fizeram parte da contracultura que influenciou a moda e os cortes de cabelo. A Inglaterra então é considerada o centro do mundo da música e da moda, homogeneizando o estilo *rock music* pelo mundo inteiro. Em cerca de 1967 a “invasão” terminou, pois as bandas estavam muito parecidas na maioria dos países.

O movimento *flower power* (poder das flores) foi um slogan usado pelos hippies dos anos 1960 até o começo dos anos 1970, representando um símbolo da ideologia da não-violência e de repúdio à Guerra do Vietnã (ANDRESEN, 2003). Uma das músicas-símbolo desse movimento é *San Francisco (Be Sure to Wear Flowers in your Hair)* (1967), que foi usada para promover o festival de rock *Monterey Pop* – a esse respeito, aliás, pode-se dizer que a década de 1960 foi a década dos festivais, sendo Woodstock, ocorrido em 1969, o mais famoso deles.

Ao mesmo tempo, no final da década de 1960, surgiu na Jamaica um gênero musical chamado reggae, que veio a ser um estilo internacional na década de 1970 e se tornou popular principalmente no Reino Unido, nos Estados Unidos e na África (ANDERSON, 2004). As bandas de reggae incorporavam estilos de muitos gêneros diferentes, incluindo *mento* (um gênero do folclore jamaicano), ska, rocksteady, calypso, soul music americano e R&B. O gênero é conhecido por sua percussão propulsiva, linhas de baixo hipnótica e estáveis, guitarra na batida chamada *skank beat*, que ajudou a marcar o ritmo como um gênero dançante. A maioria das músicas reggae possuem letras no inglês jamaicano ou em dialetos locais.

Anderson (2004) afirma que a música jamaicana possui, como assinatura, proximidade com a religião rastafari e movimentos sociais que possuem suas raízes desde a década de 1930 na Jamaica. Até hoje, muitos músicos praticam o rastafarianismo. As músicas de *reggae* frequentemente possuem temas relativos à temas espirituais rastafari (*Get up, stand up*, de Bob Marley e The Wailers, de 1973) que liga o gênero a outros estilos populares que surgiram no final da década de 1960, como o *American folk rock*. E, segundo Bradley (2000), preocupadas com certa consciência cultural, as letras de *reggae* continuaram fortes ao ponto de inspirar outros estilos musicais, tais como o hip hop.

Ainda na década de 1970, em Detroit, a maior cidade do estado de Michigan, onde existiam muitas montadoras de carros, surgiu uma gravadora de discos que trouxe o *soul* ao cenário musical. Essa gravadora chamou-se Motown e foi considerada a mais importante lançadora de artistas negros desde seu surgimento até o aparecimento do *hip hop* (GORDY, 1994). O refinamento do som da Motown pode ser considerado o precursor da era *disco*. Uma das músicas mais representativas dessa gravadora e dessa época foi *Papa was a Rollin' Stone* (1972), que, de certa maneira, resume a história da Motown e talvez dos Estados Unidos do início dos anos 1970 em sete minutos (EARLY, 2004).

Nos anos 1970, a era disco desenvolveu-se basicamente em Nova York, nos clubes de dança voltados para negros e latinos, e atingiu o auge entre 1977 e 1979, em grande parte devido ao filme *Saturday Night Fever*, de 1977. As mulheres marcaram grande presença neste estilo musical, sendo chamadas de divas, e a introdução orquestral da maioria das músicas é uma herança da Motown (ECHOLS, 2010). Uma das músicas mais representativas da era disco é *Last Dance* (1978), uma das primeiras a iniciar com uma balada, mas conter uma parte lenta no meio e depois voltar para a balada, com a frase *hook* sendo repetida várias vezes¹⁹. Porém, Echols (2010) afirma que, no começo dos anos 1980, a *disco music* começou a sofrer preconceito nos Estados Unidos, já que esse era um estilo de minorias como negros, mulheres e homossexuais.

O rock-and-roll pode ter surgido nos anos 1950 e se estabelecido na década seguinte, mas os anos 1970 foram considerados como a era do vinil, com o surgimento de artistas como David Bowie, Led Zeppelin e Pink Floyd. Também na década de 70, o movimento punk aparece como uma resposta à não-violência dos hippies e à opressão do elitismo cultural (MARCUS, 1993). Como estilo musical em si, o *punk* surgiu por volta de 1974 nos Estados Unidos, como uma manifestação cultural juvenil semelhante à das décadas de 1950 e 1960, uma espécie de retomada da cultura rock-and-roll: voltavam à cena as músicas curtas e dançantes, as jaquetas de couro estilo motociclista, camiseta branca e calça jeans, além da atitude de rebeldia. Músicas como *Anarchy in the UK* (1977), da banda Sex Pistols, eram representativas dessa época. Marcus (1993) destaca que, com exceção dos Ramones, bandas como Blondie, Talking Heads e Television estavam distantes da estética sonora e das letras que caracterizavam o *punk rock*.

Marcus (1993) relata que, enquanto o rock-and-roll tradicional ainda criava estrelas do *rock* que distanciaram o público do músico, o *punk rock* rompeu este distanciamento trazendo a música simplificada (com pouco mais de três acordes, facilmente tocados por qualquer pessoa sem formação musical) e instigando outros adolescentes a criarem suas próprias bandas. Saindo do *underground* de Nova York, nos Estados Unidos, o *punk rock* chega à Inglaterra e influencia uma série de jovens em pouco tempo (WILKINSON; WORLEY; STREET, 2017).

O *glam rock* ou *glitter rock* foi um estilo musical relacionado com uma subcultura do início dos anos de 1970, especialmente no Reino Unido. Foi uma reação contra a seriedade do rock progressivo e da contracultura do final dos anos de 1960, e

¹⁹ Tal repetição se popularizou e, até hoje, ocorre basicamente em todos os gêneros musicais.

também uma extensão desses movimentos (STRATTON, 1986). Tal estilo musical caracterizou-se por um forte apelo visual tanto dos artistas como dos seus concertos, incluindo os cabelos vivamente coloridos, os trajes escandalosos, a maquiagem pesada e até o ato de cuspir fogo (no caso do Kiss). De acordo com Stratton (1986), no *glam rock*, a música estava intimamente ligada ao desempenho do artista no palco e, a imagem do ídolo estava atrelada à criatividade dos músicos. Entre os pioneiros britânicos incluem-se David Bowie (da primeira fase) e Gary Glitter, e nos Estados Unidos, destacam-se o cenográfico Kiss e o New York Dolls (de tendência punk); nos anos de 1980, destacam-se bandas de heavy metal leve, como Bon Jovi. Entre outros grupos de *glam rock*, podem ser incluídos também o Sweet and Slade, voltado ao estilo pop prosaico, o Roxy Music e o Queen (*Bohemian Rhapsody*, 1975), mais voltados para o *art rock*. Shuker (1998) afirma que os elementos de androginia e bissexualidade eram parte da imagem e do apelo do movimento, combinando a elegância hippie à dureza *skinhead*.

Apesar de o termo *funk* surgir nos Estados Unidos na década de 1950 para descrever uma forma de jazz moderno, que se baseava no *swing* e no *soul* (sinônimos de autenticidade e sinceridade), ele passou a ter uma conotação mais negativa para se referir à música considerada grosseira ou rude. Nas décadas de 1960 e 1970, o funk foi usado para variações “anárquicas e polirrítmicas” da música soul, com “alta energia, expansão da mente pelo rock-and-roll, reação expressivamente psicodélica” (DE CURTIS; HENKE: 1992, p. 268). Entre seus principais representantes, destacam-se James Brown, George Clinton (Parliament, Funkadelic), Kool & the Gang e Earth, Wind and Fire, com músicas como *Give up the funk* (1975), da banda Parliament, considerada como uma das melhores músicas funk de todos os tempos.

O funk integrou gêneros posteriores de orientação black, como o hip-hop e o techno-funk, e o trabalho eclético de artistas como The Artist Formerly Known As Prince e Living Colour. O funk também contribuiu para a música disco e para o rap (BROWN, 1992). Na verdade, o funk engloba diversos estilos musicais e pode ser considerado um metagênero. Musicalmente, tende a ter uma pequena variação melódica, importando mais o ritmo (o “groove”):

o estilo funk requer uma formação rítmica específica – percussão e baixo – e também acordes sustentados ou interpolações rítmicas de outros instrumentos. O funk é uma atitude, que, quando expresso musicalmente

coloca o ouvinte em um estado comumente descrito como uma volta ao passado ou uma semi-embriaguez. (BROWN, 1992, p. 211)

Em 1981, o canal MTV tem o seu início com a proposta de colocar vídeos de música na televisão, ideia essa que começou na década de 1960 com os Beatles. A primeira imagem que o canal colocou no ar foi uma montagem do primeiro homem na Lua, fincando a bandeira da MTV (ROCKBIZZ, 2021) (Figura 5).

FIGURA 5 - PRIMEIRA IMAGEM QUE A MTV TRANSMITIU



Fonte: Rockbizz (2021)

A *techno music* surge em Detroit na década de 1980, como uma união da música afro-americana com temas fictícios e futuristas relevantes para a sociedade capitalista americana. Paralelamente a esse fenômeno, o *grunge* surge, principalmente em Seattle (THACKRAY, 2015). Enquanto a *techno music* tem como foco o instrumental, o *grunge* tem importância maior em suas letras, com altas doses de sarcasmo e angústia e seus músicos com uma aparência despojada para combinar com esse sarcasmo, tendo Kurt Cobain, do grupo *Nirvana*, como o representante mais emblemático desse movimento. Músicas como *Smells like teen spirit* (1991) trouxeram a estética do som aliada ao sarcasmo presente na letra.

Os anos 1980 viram Michael Jackson se reinventar e foram palco à ascensão mundial de artistas como Prince, Madonna e Whitney Houston. Michael Jackson foi o primeiro artista a ter videoclipes transmitidos pela MTV – *Beat it* e *Billie Jean*, ambos de 1982 (GEORGE, 2010) –, enquanto Donna Summer foi a primeira mulher afro-americana

a veicular vídeos na MTV. Entre os artistas que mais venderam na década estão Michael Jackson e Madonna.

Em 1990, a música *Wind of change*, da banda Scorpions, tornou-se a música tema da queda do muro de Berlim, ato que trouxe uma mudança política e comportamental na sociedade dessa década. O cenário musical viu a entrada de compositores e produtores musicais da Suécia que influenciaram a maneira como a música em língua inglesa era feita. Houve maior preocupação com a imagem dos cantores, que levou ao fenômeno do *boy band* composto por adolescentes, como Backstreet Boys, NSYNC e Westlife. Surgiram também hits de sucesso mundial, como *...Baby one more time* (1998), na voz da cantora Britney Spears (SEABROOK, 2015). No Reino Unido, o *Britpop* e o rock alternativo viram seu apogeu na década de 1990. Artistas como Coldplay, Oasis, Dido, Björk e Radiohead ficaram famosos no mundo inteiro.

O início do novo milênio foi marcado pelos atentados às torres gêmeas do *World Trade Center* em Nova York. Mas o novo milênio também trouxe a queda de preconceitos, como a eleição do primeiro presidente negro nos Estados Unidos ou a aceitação do casamento homoafetivo pela Suprema Corte americana, e mostrou novas maneiras de comunicação (mídias sociais) e de compartilhamento de conteúdo artístico, tais como YouTube, para o vídeo, e Spotify, para o áudio.

Gêneros como J-pop e K-pop surgiram na Ásia, usando uma pronúncia muito parecida com o inglês, o que foi um dos fatores para sua grande expansão e influência cultural (PALHA, 2021). A música pop indiana, por estar ligada aos filmes de Bollywood, também tornou-se popular, ao lado da música pop ocidental.

Esse cenário histórico-cultural e musical é muito importante para entendermos o período no qual o *corpus* desta pesquisa está inserido e muitos outros dados da cultura pop contemporânea podem ser encontrado em livros, artigos e na internet. Esse breve retrospecto tentou mostrar o quanto o registro música está interligado com a sociedade e a história do povo onde ele surge, mostrando também o quanto a língua inglesa está globalizada, principalmente após a Segunda Guerra Mundial. A internacionalização gerou compositores de outras nacionalidades, como o sueco Max Martin, que se apropriou tão eficientemente desse registro, a ponto de ter criado uma máquina de hits de sucesso (SEABROOK, 2015).

Como dito na Seção 1, as duas primeiras questões de pesquisa desta tese versam sobre as dimensões semânticas e acústicas, e a terceira trata da correlação canônica multimodal. A proposta desta tese é investigar elementos co-ocorrentes tanto na letra

quanto na parte acústica das músicas para entendermos melhor um registro tão presente na vida de grande parte das pessoas (muitos diariamente, como a autora) no mundo inteiro (mesmo sendo em língua inglesa), mas que ainda é tão pouco pesquisado, principalmente do ponto de vista da multimodalidade.

O próximo capítulo tratará da metodologia da tese e trará como e quais passos foram utilizados para desenvolver a pesquisa aqui relatada.

3. METODOLOGIA

*It'll be just like starting over*²⁰

John Lennon

Esta seção trata da metodologia empregada para desenvolver a pesquisa e está dividida em cinco partes. A primeira apresenta o *corpus* utilizado no estudo (Seção 3.1). A segunda, trata da análise semântica; mais especificamente, das categorias semânticas empregadas e do etiquetador USAS; a terceira, do desenvolvimento da análise multidimensional (a) semântica e (b) acústica. A última seção trata da análise de correlação canônica.

3.1 CORPORA EMPREGADOS

O corpus utilizado nesta pesquisa é chamado de CoLiE (*Corpus of Lyrics in English*)²¹. Além do corpus textual do CoLiE, que abriga a produção musical textual, foi coletada também uma base de dados (dataset) com medidas acústicas geradas pela empresa Spotify, o qual chamamos de CoLiE Spotify Dataset, ou ColiE acústico. O CoLiE abarca, portanto, dois conjuntos diferentes de dados: os intitulados, respectivamente, CoLiE *lyrics* e o CoLiE acústico, apresentados em maior detalhe a seguir.

3.1.1 Colie lyrics

O CoLiE *lyrics* foi gerado a partir de uma coleta de letras de música em língua inglesa realizada entre 1913 e 2018, e a partir de um *scraping* do site *lyrics.com* em junho de 2019. *Scraping* pode ser definido como o “processo de extração e combinação de conteúdos de interesse da internet de uma maneira sistemática”²² (GLEZ-PEÑA et al., 2014, p. 789). Cada música foi salva em um arquivo em formato texto (txt) e nomeado

²⁰ Será como começar de novo (trecho da música (just like) starting over gravada pelo cantor John Lennon em 1980).

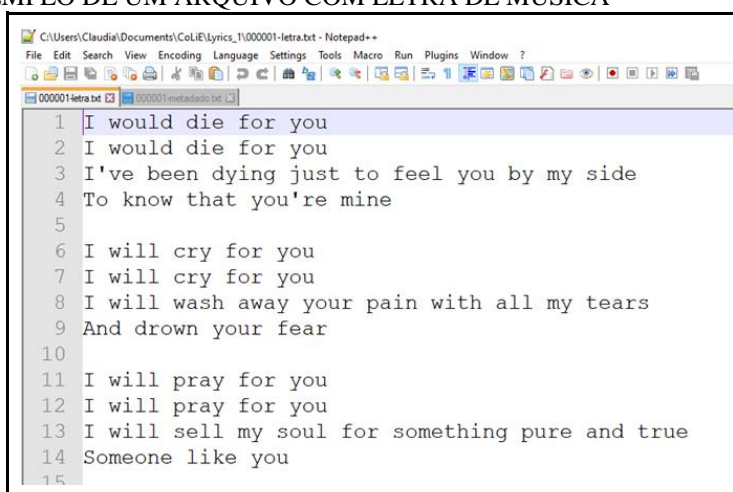
²¹ Corpus de letras de música em inglês.

²² No original: “Web data scraping can be defined as the process of extracting and combining contents of interest from the Web in a systematic way”.

numericamente, por ordem alfabética, de acordo com o título da música, por meio de um script especializado criado pelo professor orientador.

Cada arquivo do corpus corresponde a uma composição escrita, possuindo um arquivo correspondente de metadados, com informações tais como o título da música, os nomes de cantor e compositor, o ano de gravação, o gênero da música e o estilo. Tais metadados também foram coletados por *scraping* do website lyrics.com, embora o website não tenha oferecido os metadados completos para todas as composições do corpus CoLiE lyrics. As Figuras 6 e 7 ilustram o arquivo de texto 1, com a composição musical textual, e seus metadados (00001-letra e 000001-metadado), respectivamente. Esses arquivos foram produzidos por meio de um script construído pelo professor orientador.

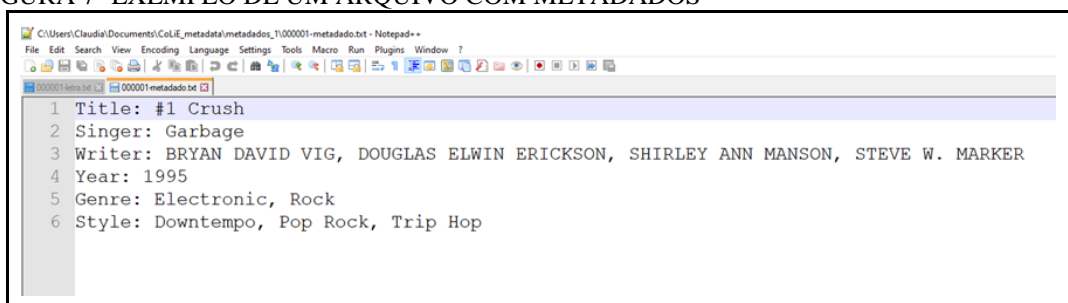
FIGURA 6 - EXEMPLO DE UM ARQUIVO COM LETRA DE MÚSICA



```
C:\Users\Claudia\Documents\CoLiE\Lyrics_1\000001-letra.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
000001-letra.txt 000001-metadado.txt
1 I would die for you
2 I would die for you
3 I've been dying just to feel you by my side
4 To know that you're mine
5
6 I will cry for you
7 I will cry for you
8 I will wash away your pain with all my tears
9 And drown your fear
10
11 I will pray for you
12 I will pray for you
13 I will sell my soul for something pure and true
14 Someone like you
15
```

Fonte: elaborada pela autora

FIGURA 7- EXEMPLO DE UM ARQUIVO COM METADADOS

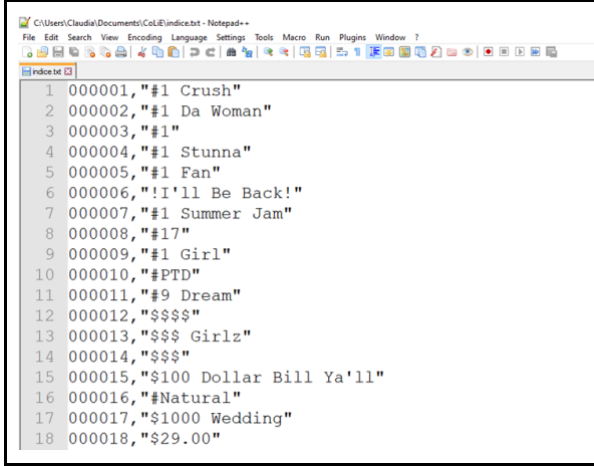


```
C:\Users\Claudia\Documents\CoLiE_metadata\metadados_1\000001-metadado.txt - Notepad++
File Edit Search View Encoding Language Settings Tools Macro Run Plugins Window ?
000001-letra.txt 000001-metadado.txt
1 Title: #1 Crush
2 Singer: Garbage
3 Writer: BRYAN DAVID VIG, DOUGLAS ELWIN ERICKSON, SHIRLEY ANN MANSON, STEVE W. MARKER
4 Year: 1995
5 Genre: Electronic, Rock
6 Style: Downtempo, Pop Rock, Trip Hop
```

Fonte: elaborada pela autora

Após os arquivos serem gerados, um outro arquivo chamado índice foi elaborado contendo o título de cada música ao lado de seu número (nome do arquivo) correspondente, como pode ser visto na Figura 8. A coleta inicial, seguindo os procedimentos acima, resultou em 236.756 arquivos individuais de letras de música.

FIGURA 8 - PARTE DO ARQUIVO ÍNDICE



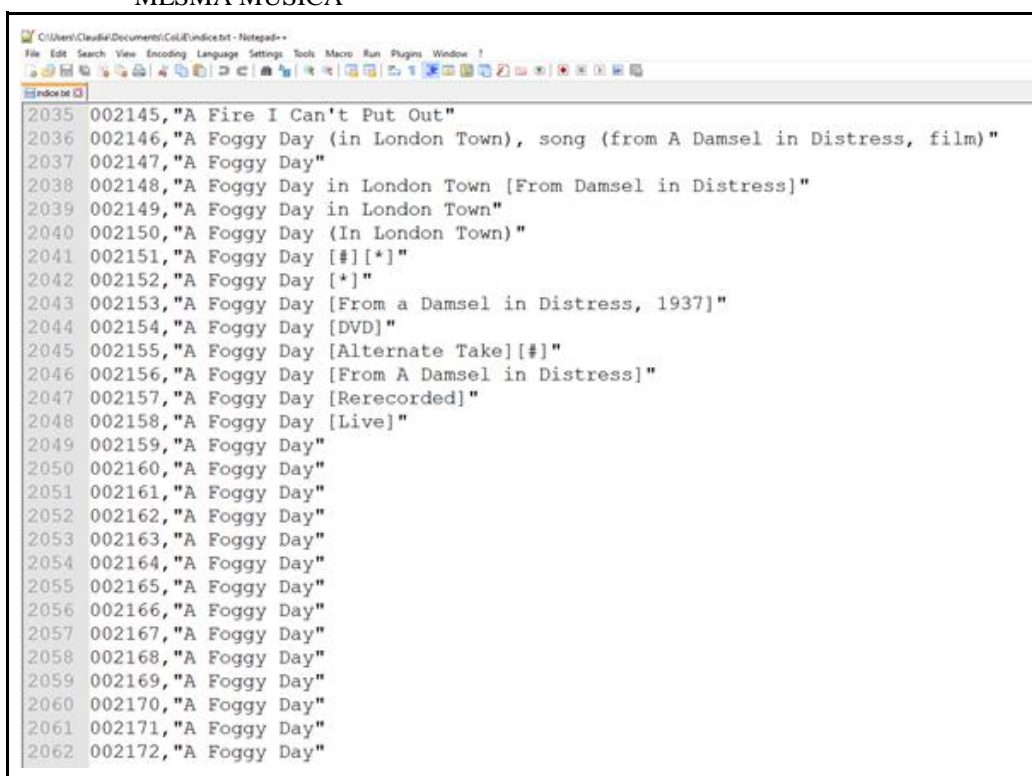
```
1 000001, "#1 Crush"  
2 000002, "#1 Da Woman"  
3 000003, "#1"  
4 000004, "#1 Stunna"  
5 000005, "#1 Fan"  
6 000006, "!I'll Be Back!"  
7 000007, "#1 Summer Jam"  
8 000008, "#17"  
9 000009, "#1 Girl"  
10 000010, "#PTD"  
11 000011, "#9 Dream"  
12 000012, "$$$$"  
13 000013, "$$$ Girlz"  
14 000014, "$$$"  
15 000015, "$100 Dollar Bill Ya'll"  
16 000016, "#Natural"  
17 000017, "$1000 Wedding"  
18 000018, "$29.00"
```

Fonte: elaborada pela autora

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.

Notamos que, quando o título da música estava em outra língua, geralmente o texto também não era em língua inglesa, então esses arquivos também foram eliminados do *corpus*. Embora intensiva, a curadoria manual foi otimizada por técnicas como o emprego do arquivo índice em ordem alfabética, o qual permitiu notar que, quando uma música era regravaada, o título era geralmente mantido – desse modo, geralmente havia uma sequência de arquivos com o mesmo nome (cf. Figura 9). A partir de uma inspeção nesses arquivos, o arquivo que correspondia à primeira gravação da música era mantido e os demais, eliminados do *corpus*.

FIGURA 9- PARTE DO ARQUIVO ÍNDICE CONTENDO ARQUIVOS DE REGRAVAÇÕES DA MESMA MÚSICA



Fonte: elaborada pela autora

A Figura 9 ilustra como exemplo as várias gravações da música *A Foggy Day*, gravada pela primeira vez por Fred Astaire em 1937, como parte do filme *A Damsel in Distress*, que corresponde ao arquivo 002146. A partir do arquivo 002147 até o 002172, o que encontramos são regravações da mesma música, sem qualquer alteração na letra. Portanto, esses arquivos foram eliminados.

Após essa inspeção e retirada dos arquivos que não correspondiam aos critérios adotados para a curadoria, o *corpus* de estudo teve a sua composição finalizada conforme ilustrado na Tabela 1. Dos 236.756 arquivos originários da coleta inicial, a curadoria eliminou mais de 30 mil arquivos.

TABELA 1 - COMPOSIÇÃO DO CORPUS COLIE LYRICS

Letras de música	Palavras únicas	Total de palavras
201.235	164.367	103.809.650

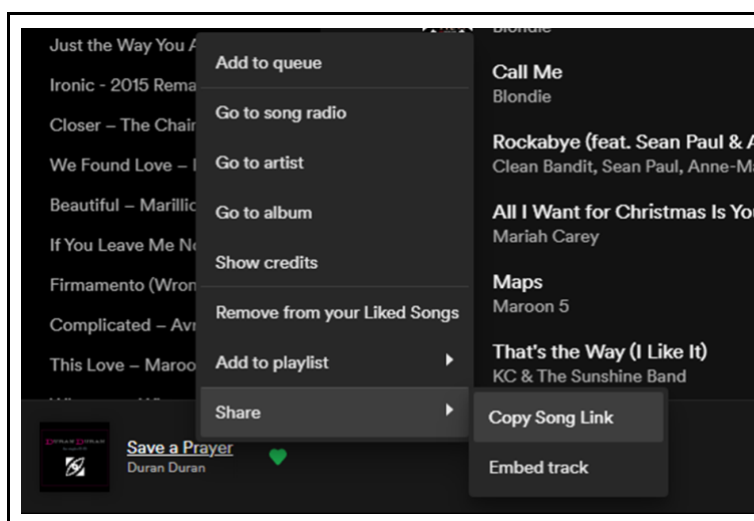
Fonte: elaborada pela autora

Uma dúvida durante a organização do corpus CoLiE *lyrics* referiu-se à necessidade ou não de transcrição das músicas, visto que algumas letras de música não correspondiam fielmente ao modo como a música fora cantada na gravação. A fim de verificar o impacto da transcrição na pesquisa, foi coletada uma amostra aleatória de 100 canções, as quais foram ouvidas e transcritas pela pesquisadora. Em seguida, as duas versões foram comparadas, a fim de verificar a extensão da diferença entre as duas versões. O resultado mostrou que, em média, havia 67 palavras únicas em comum nas duas versões, e três palavras adicionais que apareceram apenas na versão transcrita. Devido ao número baixo de palavras presentes apenas na versão transcrita, a pesquisadora optou usar as letras de música da maneira como foram coletadas do site, em vez de transcrevê-las.

3.1.2 ColiE acústico

O 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. A autora buscou o link de cada música no site de desenvolvedores do [spotify.com](https://developer.spotify.com/), conforme mostra a Figura 10. 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 fazer *scraping* dos links não forneceram bons resultados.

FIGURA 10 - COLETA DO LINK DE MÚSICAS NA PLATAFORMA SPOTIFY



Fonte: elaborada pela autora

Após os links terem sido capturados manualmente, os arquivos resultantes foram processados por um script desenvolvido pelo professor orientador, gerando uma lista, que é parcialmente mostrada na Figura 11. Algumas músicas coletadas do website lyrics.com não possuíam link na base do Spotify, como é o caso do arquivo 184590, que corresponde à música *Stone in my heart*, da banda Graffiti6, de 2010. 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.

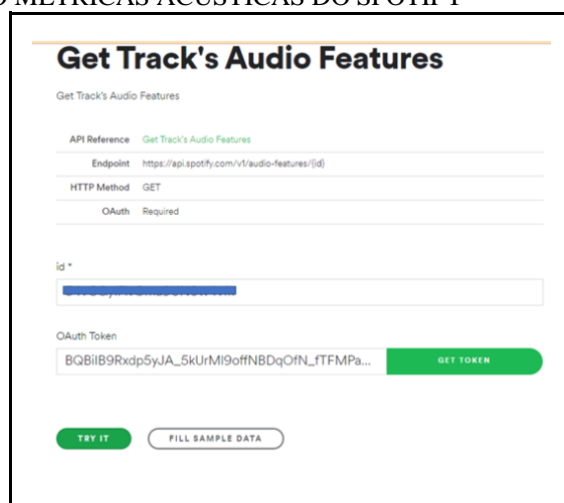
FIGURA 11 - ARQUIVO CONTENDO OS IDS DE CADA MÚSICA DA PLATAFORMA SPOTIFY

```
184577,"Stone Cold Heart" /5zQLjkuc2TuFdYE6NhmqS1?si=01e3b05b6a4d490c
184579,"Stone Believer" /4QWCVy5im8AH0WNwZ0pdj5?si=5450b02be1dc4808
184580,"Stone Cold Sober" /3BINArUJhtdRH2n176nipN?si=b48e94b6535c4a23
184581,"Stone Cold Fever" /3BWG63gwNtzGw9Qohw1XM?si=600b78cc84974a4c
184583,"Stone Cold Killers" /6PZWSvb2DB6bERLKrS99TI?si=5f5eab3bc2204af8
184584,"Stone Cold [Live][*]" /2MWfhiF38PDEM4CEnqqwPF?si=a4ab762e95a745e3
184587,"Stone Evil" /2h3ezsXx4bjhbn0qoOLVr?si=5208778ec9c94cee
184589,"Stone Glass Window" /7bHoRZIOZ5goCjTm70yo5B?si=ed55b351b3154bcc
184590,"Stone In My Heart [Stripped] [Version]" NO ID
184591,"Stone in My Shoe" /2opqo5DXdolfWtwD4rdbjB?si=1069305d5d7546bb
184592,"Stone in My Hand" /78DuwRHn1ResNuru7aJpX?si=2815f82bcc34c4a
184593,"Stone Girl" /0pfMQualHD8JdC5ZxoNZ1n?si=1a8213174a9142a8
184594,"Stone House" /1X7RG55X92aH323T2EaRRh?si=7c5d2e2bca4e4a31
```

Fonte: elaborada pela autora

Com os IDs em mãos, foi possível capturar as categorias acústicas que o API do Spotify disponibiliza²³, conforme pode ser visto nas Figuras 12 e 13. A Figura 12 mostra um passo do processo realizado: (1) colocamos o id da música previamente coletado no campo “id” da janela; (2) geramos um token em “get token”, que deve ser trocado a cada 10-15 minutos; (3) clicamos no botão “try it”.

FIGURA 12 - OBTENDO MÉTRICAS ACÚSTICAS DO SPOTIFY



Fonte: elaborada pela autora

²³ <https://developer.spotify.com/console/get-audio-features-track/>

A Figura 13 mostra as categorias que são geradas a partir dos passos tomados na Figura 12. Como pode ser visto, 11 categorias possuem valores gerados pelo API. A maioria delas possui valores entre 0 e 1, como *danceability*, *energy*, *mode*, *speechiness*, *instrumentalness*, *liveness* e *valence*. A única métrica com valores negativos é *loudness*, que fornece valores em decibéis (dB). Já a métrica *tempo* possui números elevados, diferentemente das outras métricas, por serem valores que correspondem às batidas por minuto de cada música.

FIGURA 13 - CATEGORIAS GERADAS PELO API DO SPOTIFY

```
{
  "danceability": 0.652,
  "energy": 0.662,
  "key": 5,
  "loudness": -7.926,
  "mode": 1,
  "speechiness": 0.0283,
  "acousticness": 0.203,
  "instrumentalness": 0.206,
  "liveness": 0.0977,
  "valence": 0.539,
  "tempo": 121.027,
  "type": "audio_features",
  "id": "34vCGylFlvOma9UN0w4vfk",
  "uri": "spotify:track:34vCGylFlvOma9UN0w4vfk",
  "track_href": "https://api.spotify.com/v1/tracks/34vCGylFlvOma9UN0w4vfk",
  "analysis_url": "https://api.spotify.com/v1/audio-analysis/34vCGylFlvOma9UN0w4vfk",
```

Fonte: elaborada pela autora

Um script desenvolvido pelo professor orientador também foi desenvolvido para a extração desses dados, porém alguns IDs não geraram dados e também tiveram que ser retirados, totalizando ao final 97.381 músicas compiladas no corpus acústico.

Em suma, conforme colocado anteriormente, 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.

3.2 CARACTERÍSTICAS SEMÂNTICAS

Para a consecução de estudos baseados em análise multidimensional, além de termos um *corpus* eletronicamente salvo em formato texto, é necessário que esse *corpus* esteja etiquetado. Ou seja, cada palavra precisa receber um código referente ao traço linguístico a ser analisado. Por exemplo, ao introduzir a análise multidimensional, Biber (1988) trabalhou com dados de categorias léxico-gramaticais, tais como verbo, pronome e substantivo, com o uso do etiquetador Biber Tagger. Já estudos que focam no léxico, como Kauffmann (2020), onde cada lema é um traço a ser computado, geralmente utilizam o etiquetador TreeTagger, que também já faz a lematização das palavras.

Para a análise multidimensional semântica desta pesquisa, necessitamos que campos semânticos sejam os traços linguísticos a serem analisados. Fez-se necessário utilizar um etiquetador cujas etiquetas fossem referentes a esses campos semânticos – o etiquetador USAS, desenvolvido na Universidade de Lancaster, na Inglaterra. A Figura 14 traz todas as 21 categorias principais do USAS, com suas subdivisões semânticas.

FIGURA 14 - CATEGORIAS E SUBCATEGORIAS DO ETIQUETADOR USAS

A GENERAL & ABSTRACT TERMS

- A1 General
 - A1.1.1 General actions, making etc.
 - A1.1.2 Damaging and destroying
 - A1.2 Suitability
 - A1.3 Caution
 - A1.4 Chance, luck
 - A1.5 Use
 - A1.5.1 Using
 - A1.5.2 Usefulness
 - A1.6 Physical/mental
 - A1.7 Constraint
 - A1.8 Inclusion/Exclusion
 - A1.9 Avoiding
- A2 Affect
 - A2.1 Affect: Modify, change
 - A2.2 Affect: Cause/Connected
- A3 Being
- A4 Classification
 - A4.1 Generally kinds, groups, examples
 - A4.2 Particular/general; detail
- A5 Evaluation
 - A5.1 Evaluation: Good/bad
 - A5.2 Evaluation: True/false
 - A5.3 Evaluation: Accuracy

A5.4 Evaluation: Authenticity
A6 Comparing
A6.1 Comparing: Similar/different
A6.2 Comparing: Usual/unusual
A6.3 Comparing: Variety
A7 Definite (+ modals)
A8 Seem
A9 Getting and giving; possession
A10 Open/closed; Hiding/Hidden; Finding; Showing
A11 Importance
A11.1 Importance: Important
A11.2 Importance: Noticeability
A12 Easy/difficult
A13 Degree
A13.1 Degree: Non-specific
A13.2 Degree: Maximizers
A13.3 Degree: Boosters
A13.4 Degree: Approximators
A13.5 Degree: Compromisers
A13.6 Degree: Diminishers
A13.7 Degree: Minimizers
A14 Exclusivizers/particularizers
A15 Safety/Danger

B THE BODY & THE INDIVIDUAL

B1 Anatomy and physiology
B2 Health and disease
B3 Medicines and medical treatment
B4 Cleaning and personal care
B5 Clothes and personal belongings

C ARTS & CRAFTS

C1 Arts and crafts

E EMOTIONAL ACTIONS, STATES & PROCESSES

E1 General
E2 Liking
E3 Calm/Violent/Angry
E4 Happy/sad
E4.1 Happy/sad: Happy
E4.2 Happy/sad: Contentment
E5 Fear/bravery/shock
E6 Worry, concern, confident

F FOOD & FARMING

F1 Food
F2 Drinks
F3 Cigarettes and drugs
F4 Farming & Horticulture

G GOVT. & THE PUBLIC DOMAIN

G1 Government, Politics & elections
G1.1 Government etc.
G1.2 Politics
G2 Crime, law and order
G2.1 Crime, law and order: Law & order
G2.2 General ethics
G3 Warfare, defense and the army; Weapons

H ARCHITECTURE, BUILDINGS, HOUSES & THE HOME

H1 Architecture, kinds of houses & buildings

H2 Parts of buildings

H3 Areas around or near houses

H4 Residence

H5 Furniture and household fittings

I MONEY & COMMERCE

I1 Money generally

I1.1 Money: Affluence

I1.2 Money: Debts

I1.3 Money: Price

I2 Business

I2.1 Business: Generally

I2.2 Business: Selling

I3 Work and employment

I3.1 Work and employment: Generally

I3.2 Work and employment: Professionalism

I4 Industry

K ENTERTAINMENT, SPORTS & GAMES

K1 Entertainment generally

K2 Music and related activities

K3 Recorded sound etc.

K4 Drama, the theater & show business

K5 Sports and games generally

K5.1 Sports

K5.2 Games

K6 Children's games and toys

L LIFE & LIVING THINGS

L1 Life and living things

L2 Living creatures generally

L3 Plants

M MOVEMENT, LOCATION, TRAVEL & TRANSPORT

M1 Moving, coming and going

M2 Putting, taking, pulling, pushing, transporting &c.

M3 Movement/transportation: land

M4 Movement/transportation: water

M5 Movement/transportation: air

M6 Location and direction

M7 Places

M8 Remaining/stationary

N NUMBERS & MEASUREMENT

N1 Numbers

N2 Mathematics

N3 Measurement

N3.1 Measurement: General

N3.2 Measurement: Size

N3.3 Measurement: Distance

N3.4 Measurement: Volume

N3.5 Measurement: Weight

N3.6 Measurement: Area

N3.7 Measurement: Length & height

N3.8 Measurement: Speed

N4 Linear order

N5 Quantities

N5.1 Entirety; maximum

N5.2 Exceeding; waste

N6 Frequency etc.

O SUBSTANCES, MATERIALS, OBJECTS & EQUIPMENT

O1 Substances and materials generally

O1.1 Substances and materials generally: Solid

O1.2 Substances and materials generally: Liquid

O1.3 Substances and materials generally: Gas

O2 Objects generally

O3 Electricity and electrical equipment

O4 Physical attributes

O4.1 General appearance and physical properties

O4.2 Judgment of appearance (pretty etc.)

O4.3 Color and color patterns

O4.4 Shape

O4.5 Texture

O4.6 Temperature

P EDUCATION

P1 Education in general

Q LINGUISTIC ACTIONS, STATES & PROCESSES

Q1 Communication

Q1.1 Communication in general

Q1.2 Paper documents and writing

Q1.3 Telecommunications

Q2 Speech acts

Q2.1 Speech etc: Communicative

Q2.2 Speech acts

Q3 Language, speech and grammar

Q4 The Media

Q4.1 The Media: Books

Q4.2 The Media: Newspapers etc.

Q4.3 The Media: TV, Radio & Cinema

S SOCIAL ACTIONS, STATES & PROCESSES

S1 Social actions, states & processes

S1.1 Social actions, states & processes

S1.1.1 General

S1.1.2 Reciprocity

S1.1.3 Participation

S1.1.4 Deserve etc.

S1.2 Personality traits

S1.2.1 Approachability and Friendliness

S1.2.2 Avarice

S1.2.3 Egoism

S1.2.4 Politeness

S1.2.5 Toughness; strong/weak

S1.2.6 Sensible

S2 People

S2.1 People: Female

S2.2 People: Male

S3 Relationship

S3.1 Relationship: General

S3.2 Relationship: Intimate/sexual

S4 Kin

S5 Groups and affiliation

S6 Obligation and necessity

S7 Power relationship
S7.1 Power, organizing
S7.2 Respect
S7.3 Competition
S7.4 Permission
S8 Helping/hindering
S9 Religion and the supernatural

T TIME

T1 Time
T1.1 Time: General
T1.1.1 Time: General: Past
T1.1.2 Time: General: Present; simultaneous
T1.1.3 Time: General: Future
T1.2 Time: Momentary
T1.3 Time: Period
T2 Time: Beginning and ending
T3 Time: Old, new and young; age
T4 Time: Early/late

W THE WORLD & OUR ENVIRONMENT

W1 The universe
W2 Light
W3 Geographical terms
W4 Weather
W5 Green issues

X PSYCHOLOGICAL ACTIONS, STATES & PROCESSES

X1 General
X2 Mental actions and processes
X2.1 Thought, belief
X2.2 Knowledge
X2.3 Learn
X2.4 Investigate, examine, test, search
X2.5 Understand
X2.6 Expect
X3 Sensory
X3.1 Sensory: Taste
X3.2 Sensory: Sound
X3.3 Sensory: Touch
X3.4 Sensory: Sight
X3.5 Sensory: Smell
X4 Mental object
X4.1 Mental object: Conceptual object
X4.2 Mental object: Means, method
X5 Attention
X5.1 Attention
X5.2 Interest/boredom/excited/energetic
X6 Deciding
X7 Wanting; planning; choosing
X8 Trying
X9 Ability
X9.1 Ability: Ability, intelligence
X9.2 Ability: Success and failure
Y SCIENCE & TECHNOLOGY
Y1 Science and technology in general
Y2 Information technology and computing

Z NAMES & GRAMMATICAL WORDS

Z0 Unmatched proper noun
Z1 Personal names
Z2 Geographical names
Z3 Other proper names
Z4 Discourse Bin
Z5 Grammatical bin
Z6 Negative
Z7 If
Z8 Pronouns etc.
Z9 Trash can
Z99 Unmatched

Fonte: Archer et al. (2002)

Como pode ser visto na Figura 14, as categorias semânticas são compostas de:

1. Uma letra maiúscula, indicando o campo semântico principal;
2. Um número, indicando uma primeira subdivisão do campo semântico;
3. Um ponto decimal seguido de um número, indicando uma subdivisão mais refinada do campo semântico (que não está presente em todos);
4. Um ou mais sinais de mais (+) ou menos (-), indicando uma posição semântica positiva ou negativa na escala semântica (que não está presente em todos);
5. Uma barra seguida de uma segunda etiqueta, indicando dúvida entre duas categorias (que não está presente em todos);
6. As letras “f” ou “m”, indicando respectivamente os gêneros feminino (f) e masculino (m), se aplicável.

3.3 ANOTAÇÃO SEMÂNTICA

O corpus CoLiE *lyrics* foi etiquetado com o etiquetador semântico USAS, conforme visto acima. O site <http://ucrel-api.lancaster.ac.uk/usas/tagger.html>, que é capaz de fazer a etiquetagem automática de textos com até 100.000 palavras, foi insuficiente para realizar a tarefa em relação ao corpus da presente pesquisa, que contém 103.809.650 palavras. Para realizar o processo de etiquetagem de maneira mais rápida, foi utilizado como alternativa um script em linguagem Python, desenvolvido pelo professor Joe Collentine, de modo a etiquetar todo o *corpus*.

Através de outro script, também desenvolvido em Python pelo professor Joe Collentine, os textos foram colocados em uma planilha, onde cada campo semântico tornou-se uma variável e os valores de frequência de cada campo semântico foram computados, conforme mostra a Figura 15.

FIGURA 15 - PLANILHA COM OS CAMPOS SEMÂNTICOS DO *CORPUS COLIE*

file	A1	A2	A3	A4	A5	A6	A7	A8	A9	B'
Lyrics_29-224879-letra.txt		13	2	11	0	4	3	2	2	1
Lyrics_15-118457-letra.txt		4	0	0	0	1	0	0	0	4
Lyrics_9-070975-letra.txt		8	7	17	0	4	3	1	0	5
Lyrics_10-078943-letra.txt		21	1	10	0	6	0	3	0	2
Lyrics_5-033186-letra.txt		2	0	26	0	4	0	1	0	3
Lyrics_29-228099-letra.txt		3	0	13	2	1	0	3	0	0
Lyrics_26-200866-letra.txt		25	2	22	2	4	0	3	2	2
Lyrics_25-195964-letra.txt		15	0	9	0	0	0	0	0	6
Lyrics_7-055958-letra.txt		28	2	14	9	7	1	5	1	18
Lyrics_18-137027-letra.txt		12	0	3	0	0	2	0	0	2
Lyrics_16-121084-letra.txt		7	2	11	0	1	0	3	0	3
Lyrics_20-153713-letra.txt		19	4	13	0	0	3	3	0	19
Lyrics_25-199184-letra.txt		2	0	4	0	0	2	1	0	4
Lyrics_17-132495-letra.txt		12	4	18	0	0	2	5	0	5
Lyrics_6-043441-letra.txt		14	0	6	1	1	0	5	0	6
Lyrics_7-050050-letra.txt		10	4	9	1	2	1	4	0	5
Lyrics_25-199467-letra.txt		14	0	15	0	0	0	2	0	2
Lyrics_27-213994-letra.txt		15	0	21	0	1	0	5	0	4
Lyrics_24-186896-letra.txt		7	0	3	0	0	0	1	0	0
Lyrics_17-132176-letra.txt		19	0	9	0	1	0	8	0	7
Lyrics_10-074440-letra.txt		0	0	4	0	2	0	3	0	0
Lyrics_11-081556-letra.txt		0	0	19	0	0	0	0	0	0
Lyrics_22-170639-letra.txt		5	2	8	8	2	1	8	0	21
Lyrics_23-178200-letra.txt		6	1	7	0	2	2	8	0	1
Lyrics_12-090123-letra.txt		6	0	12	0	2	0	0	0	0
Lyrics_8-063887-letra.txt		9	0	7	3	1	0	4	0	4
Lyrics_15-114954-letra.txt		6	2	11	0	1	0	1	0	2
Lyrics_8-059928-letra.txt		6	0	6	0	1	0	0	0	1

Fonte: elaborada pela autora

3.4 ANÁLISE MULTIDIMENSIONAL SEMÂNTICA

A análise fatorial é o procedimento estatístico usado na análise multidimensional para reduzir o conjunto inicial de variáveis (traços linguísticos) para um número menor de fatores, que, juntos, atuam como “parâmetros linguísticos subjacentes” (EGBERT; BIBER, 2018, p. 2) na medida em que são responsáveis por grande parte da variação encontrada. Na vertente semântica, cada dimensão da AMD corresponde a determinado agrupamento semântico existente no registro em estudo (letras de música em língua inglesa), que é interpretado pela pesquisadora a partir do efeito da frequência associada das variáveis coocorrentes das dimensões sobre os textos do *corpus*. Já na AMD acústica, a dimensão resultante da análise fatorial corresponde a aspectos acústicos presentes nas músicas do *corpus*. As dimensões são nomeadas também a partir de grupos coocorrentes de variáveis acústicas, formadas nesta pesquisa pelas onze categorias elencadas pela

plataforma Spotify. Em ambas as AMDs, as dimensões geradas pela análise fatorial expressam variáveis latentes que atuam indiretamente na variação encontrada na superfície, interpretadas a partir de variáveis coocorrentes nos fatores.

Na análise fatorial, importa, portanto, reconhecer quais variáveis coocorrem a partir de uma matriz de correlação (KAUFFMANN, 2020). Dada a natureza exploratória da pesquisa, foi utilizado um método de análise fatorial por eixos principais (*principal axis factoring*, ou PAF), que tem como critério priorizar as variáveis com maior covariância para a formação dos fatores. Tal metodologia, ademais, segue o padrão utilizado na maior parte dos estudos de AMD, como Biber (1988), Berber Sardinha, Kauffmann e Acunzo (2014) e Kauffmann (2020).

A análise fatorial requer uma avaliação prévia de adequação e conformidade dos dados, antes da efetiva análise: a boa prática (FRIGINAL; HARDY, 2014; LOEWEN; GONULAL, 2015; KAUFFMANN, 2020) recomenda efetuar exames para verificação de multicolinearidade e de baixa comunalidade nas variáveis, para que a extração fatorial final seja composta de variáveis que não estejam *a priori* correlacionadas (multicolinearidade), ou que contribuam para a covariação geral do conjunto (comunalidade maior que 0,15). Multicolinearidade é o efeito causado por variáveis que apresentam alta correlação entre si (coeficiente de relação de Pearson acima de 0,8). Quando duas variáveis estão nesta situação de sobreposição – quase de singularidade, quando toda covariação é coincidente –, é preciso retirar uma das variáveis para evitar que a extração fatorial seja influenciada por essa relação direta entre as variáveis. Para identificar essas variáveis é preciso gerar uma matriz de correlação de Pearson e verificar os maiores valores (função RANK do comando PROC CORR, no programa SAS). Enquanto o exame da multicolinearidade antecede a análise fatorial, o exame das comunalidades é realizado depois de algumas extrações fatoriais intermediárias, nas quais se experimentam algumas composições de variáveis em diversos níveis de granularidade, até a extração fatorial final não rotacionada, pois suas estimativas dependem do rol de variáveis examinado no momento da pesquisa (KAUFFMANN, 2020).

Além da eliminação das variáveis com muita ou pouca colinearidade, faz-se necessária a escolha das variáveis que serão utilizadas na análise fatorial. Para estudos lexicais, geralmente essa escolha é feita a partir das palavras com frequência destacada no *corpus*. Esse procedimento não funcionou no *corpus* CoLiE *lyrics*, pois o resultado da análise mostrava fatores com apenas uma variável. Um outro método de escolha das variáveis teve então que ser considerado para esse *corpus* e, para tanto, utilizamos o

procedimento VARCLUS do SAS, que distribui as variáveis em grupos a partir de uma medida de similaridade, onde as variáveis do mesmo cluster possuem uma distribuição semelhante entre os textos.

Com as variáveis determinadas para a análise fatorial, partiu-se para outro procedimento importante, que vem a ser o exame do gráfico de sedimentação, ou *scree plot*, baseado na variância dos autovalores iniciais. No momento da extração fatorial final não rotacionada, esse gráfico projeta visualmente a carga de comunalidade alcançada em cada autovalor (ou *eigenvalue*), em ordem decrescente, até que o último autovalor, no total das cargas de comunalidade somada dos autovalores, representem 100% da variação apresentada nos dados. A análise do *scree plot* se concentra nos primeiros autovalores, porque serão esses os maiores responsáveis pela variação do conjunto: o primeiro autovalor tem a maior importância, pois explica a maior quantidade de variação; o segundo autovalor tem capacidade de explicar uma quantidade de variação menor que o primeiro autovalor, e assim por diante.

O número de fatores a ser escolhido na extração fatorial final rotacionada é definido basicamente pela análise do gráfico de sedimentação com esses autovalores iniciais. O método para determinar o número de fatores é examinar no gráfico de sedimentação onde ocorre a primeira quebra na descida dos primeiros autovalores, onde a linha parece ficar mais plana; o número de autovalores que estiver no gráfico à esquerda dessa quebra normalmente define o número de fatores na extração fatorial final, podendo variar um pouco a mais ou a menos (KAUFFMANN, 2020).

O analista deve utilizar um pouco de subjetividade ao aplicar a regra para buscar a solução mais interpretável entre as alternativas, e optar pela que mais se aproximar desse parâmetro. É preciso considerar também que a solução não é ainda a definitiva, já que, de posse da informação do número de fatores, é feita a extração final rotacionada, que apresenta uma outra combinação de agrupamento de variáveis nos fatores, em geral em uma configuração mais amigável para a interpretação. Nas análises multidimensionais realizadas nesta pesquisa, foi utilizado o método de rotação oblíqua Promax, mais utilizado nos estudos de AMD (BIBER, 1988; BIBER et al., 2006; BERBER SARDINHA; KAUFFMANN; ACUNZO, 2014; KAUFFMANN, 2020) porque leva em consideração as possíveis intercorrelações entre fatores, o que é algo de se esperar em se tratando de dados linguísticos interdependentes.

Após a definição do número de fatores, são computados escores individuais de observações por fator e também escores médios, formados pelos escores de frequência

das observações do corpus.

Isso posto, podemos resumidamente dizer que uma análise multidimensional divide-se basicamente em três partes principais, que serão destacadas a seguir, com as especificidades do corpus CoLiE *lyrics* utilizado nesta pesquisa.

(1) Pré-processamento do *corpus*:

- Etiquetagem dos textos efetuada por meio do etiquetador USAS;
- Normalização dos textos por 100 palavras²⁴, de modo que textos de tamanhos diferentes consigam ser comparados sem que um texto com muitas palavras sobressaia em relação a um texto com menos palavras;
 - Seleção das variáveis semânticas realizada a partir da análise de clusters, que revelou quais variáveis estavam mais correlacionadas entre si. O programa utilizado para esta análise de clusters das variáveis no SAS foi desenvolvido pelo professor orientador, e está reproduzido no Anexo 1.

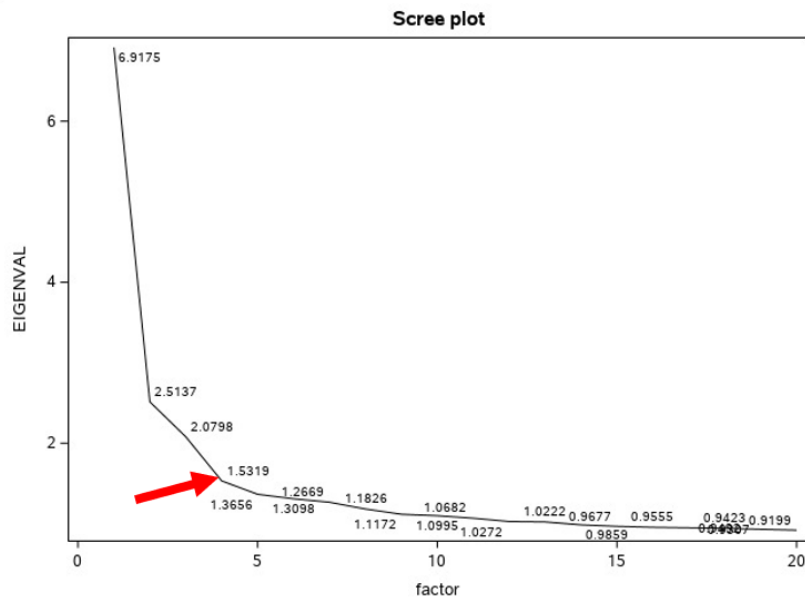
(2) Primeira análise fatorial (não rotacionada):

- Extração da primeira análise fatorial pelo programa *SAS on Demand for Academics*²⁵, no qual determinou-se em quatro o número final de fatores, conforme indica a seta na Figura 16.

²⁴ Geralmente a normalização de textos é feita por 1000, porém recomenda-se que textos pequenos como letras de música, tweets, postagens de redes social em geral, tenham sua normalização feita por 100 ao invés de por 1000, como na presente pesquisa.

²⁵ <https://welcome.oda.sas.com/login> (Aceso em 30 de janeiro de 2022)

FIGURA 16 - GRÁFICO DE ESCARPA DA ANÁLISE SEMÂNTICA



Fonte: elaborada pela autora

(3) Segunda análise fatorial (rotacionada):

- Retirada de variáveis com peso muito baixo na solução rotacionada;
- Extração final da segunda análise fatorial, na qual as variáveis foram distribuídas em cada um dos quatro fatores, de acordo com sua coocorrência. O programa SAS utilizado para processar a análise fatorial foi desenvolvido pelo professor orientador e está representado no Apêndice 2. A Tabela 2 mostra todas as variáveis presentes em cada fator, em ordem decrescente de peso da variável no fator;
- Cálculo dos escores dos textos e escores médios dos fatores;
- Realização de uma análise qualitativa dos textos, levando-se em conta os resultados dos escores de cada variável em cada fator, na qual são nomeadas as dimensões semânticas.

TABELA 2 - PADRÃO FATORIAL DA ANÁLISE SEMÂNTICA

Fator	Variável	Peso
F1	F1 - food	0,62
	B5 - clothes and belongings	0,57
	I2 - business	0,56
	L2 - living creatures: animals, birds, etc.	0,54
	F2 - drinks and alcohol	0,50
	K5 - sports and games generally	0,48
	I1 - money generally	0,47
	Q4 - the media	0,45
	M3 - vehicles and transport on land	0,43
	O3 - electricity and electrical equipment	0,41
	P1 - education in general	0,40
	H3 - areas around or near the house	0,38
	M4 - sailing, swimming, etc.	0,37
	S2 - people	0,37
	Y2 - information technology and computing	0,36
	Y1 - science and technology in general	0,36
	H1 - architecture, houses and buildings	0,35
K4 - drama, the theater and show business	0,32	
Q1 - linguistics actions, states and processes; communication	0,31	
F2	Q3 - language, speech and grammar	0,34
	X1 - psychological actions, states and processes	0,32
	X8 - trying	0,30
F3	W3 - geographical terms	0,52
	W1 - the universe	0,46
	W2 - light	0,41
	M5 - flying and aircraft	0,31
	S9 - religion and supernatural	0,31
F4	M1 - moving, coming and going	0,31
	H4 - residence	0,30

Fonte: elaborada pela autora

Ao inspecionar a Tabela 2, vemos que todos os fatores possuem apenas um polo (positivo). O Fator 1 possui a maior quantidade de variáveis e, portanto, concentra a maior variação do corpus. No Capítulo 4, pode ser visto como esses resultados são interpretados.

Após apresentarmos no próximo tópico as características acústicas, traremos a AMD acústica em maior detalhe.

3.5 CARACTERÍSTICAS ACÚSTICAS

Panda et al. (2021) resume as características acústicas do aplicativo Spotify na Tabela 3, abaixo. As variáveis utilizadas pelo referido autor coincidem com as utilizadas na presente pesquisa.

TABELA 3 - VARIÁVEIS ACÚSTICAS DO SPOTIFY

Variável	Significado
Musicalidade	Medida de confiança que vai de 0 a 1 e mostra se a faixa é ou não acústica. Valores próximos a 1,0 representam alta confiança de que a faixa é acústica.
Dançabilidade	Descreve o quão adequada uma faixa é para dançar, baseado numa combinação de elementos incluindo tempo, estabilidade do ritmo, força da batida e regularidade, de maneira geral. 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 <i>loudness</i> , timbre e entropia, de maneira geral.
Instrumentalidade	Prevê se uma faixa contém ou não vocais. Sons tais como “Ooh” e “aah” são tratados como instrumentais. Quanto mais próximo o valor estiver de 1, maior a chance da faixa ser instrumental. Valores acima de 0,5 tendem a representar faixas instrumentais.
Tom	Tom geral da faixa.
Vivacidade	Detecta a presença de público na gravação. Quanto mais alto os valores de <i>liveness</i> , maior a probabilidade de que a faixa seja ao vivo. Um valor acima de 0,8 fornece grande probabilidade de que a faixa seja ao vivo.
Volume	Valor geral de uma faixa em decibéis (dB). Os valores de <i>loudness</i> são medidos por toda a faixa e podem comparar o volume das faixas. Valores típicos de <i>loudness</i> variam de -60 a 0dB.
Modalidade	Indica a modalidade (maior ou menor) de uma faixa, o tipo de escala do qual o conteúdo melódico é derivado. Maior modalidade é representado por 1, e menor por 0 (apenas esses dois valores são elencados).
Verbosidade	Detecta a presença de palavras faladas em uma faixa. Quanto mais a faixa contém verbosidade (por exemplo, talk show, audio book e poesia), mais próximo de 1 é o valor. Valores acima de 0,66 descrevem faixas que provavelmente são feitas totalmente de palavras faladas. Valores entre 0,33 e 0,66 descrevem faixas que podem conter tanto música como fala. Valores abaixo de 0,33 representam música e outras faixas sem fala.
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.

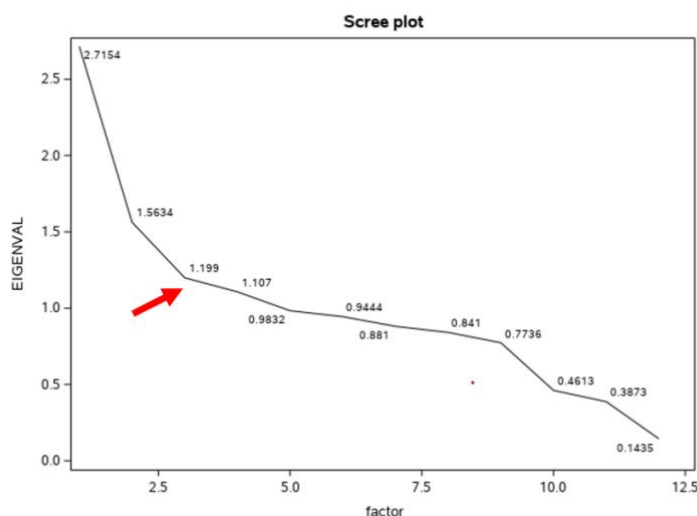
Valência	Medida numérica entre 0 e 1 que descreve a positividade musical que uma faixa possui. Quanto mais alto o valor, mais positiva a música é, enquanto que valores mais baixos representam músicas negativas.
Compasso	Especifica quantas batidas existem em cada medida, variando de 3 a 7 (3/4, 4/4, 5/4, 6/4, 7/4), em notas de 1/4.

Fonte: adaptada de Panda et al. (2021)

3.6 ANÁLISE MULTIDIMENSIONAL ACÚSTICA

A análise multidimensional acústica foi realizada no corpus CoLiE acústico utilizando as onze categorias descritas acima. O gráfico de escarpa gerado pela análise fatorial não rotacionada inicial revelou a presença de três fatores, conforme pode ser visto na Figura 17.

FIGURA 17 - GRÁFICO DE ESCARPA DA ANÁLISE ACÚSTICA



Fonte: elaborada pela autora

Uma segunda análise fatorial rotacionada foi executada, e os escores foram calculados com base nas características acústicas que carregaram em cada fator. O programa SAS utilizado para esta análise fatorial foi desenvolvido pelo professor orientador e está representado no Apêndice 3. A Tabela 4 mostra as variáveis presentes em cada fator, em ordem decrescente de peso da variável no fator.

TABELA 4 - PADRÃO FATORIAL DA ANÁLISE ACÚSTICA

Fator	Variável	Peso
F1+	energia	0,85
	volume	0,79
	ritmo	0,25
	vivacidade	0,16
F1-	musicalidade	-0,74
F2+	dançabilidade	0,69
	valência	0,65
	compasso	0,24
	verbosidade	0,23
F2-	instrumentalidade	-0,28
F3+	modalidade	0,38
F3-	tom	-0,29

Fonte: elaborada pela autora

Diferentemente da análise multidimensional semântica, todos os três fatores da análise acústica apresentam dois pólos (positivo e negativo); conforme colocado, os termos “positivo” e “negativo” não retratam julgamento de valor, apenas indicam que as características acústicas de cada polo estão em distribuição complementar – isto é, quando as de um pólo estão em evidência em uma canção, as do polo oposto não estão. Por exemplo, no polo positivo do Fator 1 carregaram músicas que tendem a possuir maior peso nas variáveis energia, volume, ritmo e vivacidade e quase não apresentam músicas com alto peso em musicalidade; por outro lado, no polo negativo do mesmo Fator 1, carregaram músicas que apresentam um alto peso em musicalidade e que quase não apresentam músicas com alto peso nas variáveis energia, volume, ritmo e vivacidade, conforme pode ser observado na Figura 18.

FIGURA 18 - POLOS DO FATOR 1 DA ANÁLISE MULTIDIMENSIONAL ACÚSTICA



Fonte: elaborado pela autora

3.7 ANÁLISE DE CORRELAÇÃO CANÔNICA

A análise canônica incumbe-se de verificar se há, e de que maneira ocorre, uma relação entre dois conjuntos de variáveis (correlação) que mensuram diferentes aspectos de um mesmo conjunto de observações – neste caso, os escores de cada música do Corpus CoLie acústico – atribuídos pelas análises multidimensionais semântica e acústica realizadas nesta pesquisa. Segundo Tabachnick e Fidell (2013, p. 262):

O objetivo da correlação canônica é analisar a relação entre dois grupos de variáveis. A correlação canônica fornece uma análise estatística onde cada observação é medida em dois grupos de variáveis e o pesquisador quer saber se e como os dois grupos se relacionam.²⁶

A utilidade do procedimento estatístico da análise canônica, apesar de implicar um grau maior de complexidade no modelo de interpretação dos resultados, confere à pesquisa a chance de explorar as relações existentes entre diferentes pares de análises multidimensionais efetuadas previamente, no caso da presente pesquisa as AMDs semântica e acústica do corpus CoLiE. Porém, de acordo com Kauffmann (2020), a análise canônica não permite que seu resultado seja entendido como uma relação de causalidade entre variáveis de dois conjuntos, mas apenas como uma relação de associação, assim como toda correlação. Portanto, a presente pesquisa limita-se somente a observar a emergência de pares canônicos, relacionando dois subconjuntos de variáveis

²⁶No original: “The goal of canonical correlation is to analyze the relationships between two sets of variables. Canonical correlation provides a statistical analysis for research in which each subject is measured on two sets of variables and the researcher wants to know if and how the two sets relate to each other.”

correlacionadas, de modo a permitir uma visão global da dinâmica de dimensões obtidas sob diferentes prismas, ou mensurações de sistemas de variáveis dependentes, de tal forma que seja possível enriquecer a discussão sobre a relação intrínseca entre a letra e a melodia da música tratada neste trabalho.

A partir dos dois conjuntos de variáveis de dimensões produzidas pelas AMDs²⁷, foram calculadas na rotina CANCECORR, do programa SAS, as correlações lineares entre as variáveis do primeiro e do segundo conjunto de variáveis, separadamente, e mais duas matrizes de correlação entre os dois conjuntos. Essas matrizes produziram autovalores e autovetores (*eigenvalues* e *eigenvectors*) que atribuíram valores compartilhados de variância entre os dois conjuntos às cargas de variância originais, resultando em correlações canônicas das variáveis de um conjunto, combinadas em pares com as correlações canônicas das variáveis do outro conjunto, produzindo pares canônicos.

Uma análise canônica pode ter mais de um par canônico significativo, e a ordem dos pares canônicos resultantes da análise canônica também obedece a uma ordem decrescente em magnitude, assim como ocorre na análise fatorial. Cada par canônico significativo é considerado independente dos demais pares canônicos, uma vez que a solução da análise canônica é ortogonal (TABACHNICK; FIDELL, 2013). A ausência de coeficientes canônicos maiores que 0,30 indica que o par canônico é de importância secundária e é descartado por não ser, em geral, interpretável (KAUFFMANN, 2020).

Na presente pesquisa, foi realizada uma análise canônica como produto da combinação de AMDs semântica e acústica. Para tanto, o *corpus* utilizado foi o CoLiE acústico, com 95.913 casos, pois os arquivos de texto e de áudio devem ser os mesmos para a análise canônica poder ser processada.

Entre os dados analisados nos resultados da análise canônica estão os coeficientes das variáveis canônicas ajustadas, as correlações canônicas quadradas, que mostra qual é a variância compartilhada entre os dois conjuntos de variáveis, no tocante à variável canônica, e os testes da hipótese nula H_0 . Entre esses testes foram calculados os valores aproximados de F , relatando a força da associação entre conjuntos, o numerador e o denominador dos graus de liberdade e o valor de p (coluna “Pr>F”), a fim

²⁷ Vale ressaltar que foram utilizados na análise canônica 95.913 arquivos que estavam presentes ao mesmo tempo nas dimensões semânticas (CoLiE lyrics) e nas dimensões acústicas (CoLiE acústico).

de verificar a chance de a variável canônica ser produzida por obra do acaso, no limite de significância fixado em 0,05. Conforme a Tabela 5 mostra, as três dimensões canônicas possuem valor de p menor do que 0,05.

TABELA 5 - TESTE DA HIPÓTESE NULA H0

Razão de Verossimilhança (Likelihood Ratio)	Valor Aproximado de F	Num DF	Den DF	Pr>F
0,90	824,63	12	2584588	<0,0001
0,99	30,52	6	195476	<0,0001
0,99	6,05	2	97739	0,0024

Fonte: elaborada pela autora

Mais testes complementares verificaram a estabilidade da solução, como o teste de Wilks, o traço de Pillai, o traço de Hotelling-Lawley e a raiz maior de Roy. De acordo com a Tabela 6 podemos ver que o valor de p foi <0,05 (significativo) em todos os testes.

TABELA 6 - TESTES COMPLEMENTARES PARA ATESTAR A CONFIABILIDADE DA ANÁLISE DE CORRELAÇÃO CANÔNICA

Estatística multivariada e aproximações de F					
S=3 M=0 N=48867,5					
Estatística	Valor	Valor de F	Num DF	Den DF	Pr>F
Hipótese de Wilks	0,90	824,63	12	258588	<.0001
Traço de Pillai	0,09	796,95	12	293217	<.0001
Traço de Hotelling-Lawley	0,10	849,25	12	171035	<.0001
Maior raiz de Roy	0,10	2502,01	4	97739	<.0001

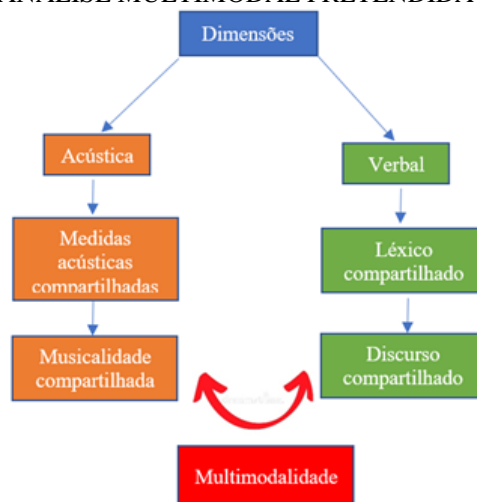
Nota: Estatística F para a Maior raiz de Roy é um limite superior

Fonte: elaborada pela autora

O interesse na exploração dos dados multivariados que contribuem para a variação na música é, afinal, o intuito de incluir na pesquisa a análise canônica, um procedimento estatístico recente na pesquisa multidimensional multimodal.

Pode-se dizer que a análise de correlação canônica fornece o caminho para a análise da multimodalidade, conforme ilustrado na Figura 19:

FIGURA 19 – SÍNTESE DA ANÁLISE MULTIMODAL PRETENDIDA NO TRABALHO



Fonte: elaborada pela autora

Com base na Figura 19, podemos ver que as dimensões acústicas e semânticas revelam as características compartilhadas que, quando unidas através da análise canônica, podem revelar a multimodalidade do registro música em língua inglesa, na medida em que representam dois modos semióticos distintos, o verbal (capturado pela análise multidimensional semântica) e o acústico (capturado pela análise multidimensional acústica).

4. RESULTADOS E DISCUSSÃO

*Like Frankie said, 'I did it my way'*²⁸
Martin Gellner / Werner Stranka

Nesta seção apresentamos os resultados da pesquisa, começando com a interpretação dos fatores descritos no capítulo anterior em dimensões semânticas e acústicas. A seguir são descritas as dimensões multimodais, resultado da análise canônica que correlaciona as dimensões semânticas e acústicas.

4.1 RESULTADOS DA AMD SEMÂNTICA

A partir dos resultados da análise fatorial, chegamos à Tabela 2 utilizada na metodologia e reproduzida novamente, na sequência. Tal tabela foi organizada em ordem decrescente de pesos das variáveis que carregaram em cada fator. Para melhor entendimento e facilitação da interpretação, apresentaremos a função que cada variável apresentou nos textos inspecionados na análise qualitativa.

TABELA 2 - PADRÃO FATORIAL DA ANÁLISE SEMÂNTICA

Fator	Variável	Peso
F1	F1 - food	0,62
	B5 - clothes and belongings	0,57
	I2 - business	0,56
	L2 - living creatures: animals, birds, etc.	0,54
	F2 - drinks and alcohol	0,50
	K5 - sports and games generally	0,48
	I1 - money generally	0,47
	Q4 - the media	0,45
	M3 - vehicles and transport on land	0,43
	O3 - electricity and electrical equipment	0,41
	P1 - education in general	0,40
	H3 - areas around or near the house	0,38
	M4 - sailing, swimming, etc.	0,37
	S2 - people	0,37
	Y2 - information technology and computing	0,36
Y1 - science and technology in general	0,36	

²⁸ Como Frankie (Sinatra) disse, “eu fiz do meu jeito” (trecho da música *It's my life*, gravada pela banda Bon Jovi).

	H1 – architecture, houses and buildings K4 – drama, the theater and show business Q1 – linguistics actions, states and processes; communication	0,35 0,32 0,31
F2	Q3 - language, speech and grammar X1 – psychological actions, states and processes X8 – trying	0,34 0,32 0,30
F3	W3 - geographical terms W1 - the universe W2 - light M5 - flying and aircraft S9 - religion and supernatural	0,52 0,46 0,41 0,31 0,31
F4	M1 - moving, coming and going H4 - residence	0,31 0,30

Fonte: elaborada pela autora

4.1.1 - Fator 1, análise semântica

O Fator 1 possui dezenove variáveis. Como a maioria dos estudos de análise fatorial, o primeiro fator abarca mais variáveis e, portanto, concentra a maior porcentagem de variação do corpus (6,92%). Essas variáveis possuem funções específicas reveladas a partir da análise qualitativa, que envolveu a inspeção de uma amostra dos textos que mais carregaram neste fator. A seguir, explicaremos cada uma dessas funções, com exemplos do *corpus* para facilitar a compreensão.

A primeira variável é F1 “food”, que não se refere apenas a comida de forma literal (*Evaporated meats* - Exemplo 1), mas também a uma maneira carinhosa de tratar a pessoa com quem o artista está simulando uma conversa (*Honey, it's a royal treatment no tracks out* - Exemplo 2). Na sequência, temos a variável B5 “clothes and personal belongings”, que marcadamente mostra uma ostentação de marcas de roupa, carros, aparência – uma crítica à exaltação da riqueza, da mesma maneira que a variável I1 “money generally” (*Expensive jeans* - Exemplo 1). A variável I1 também aparece fortemente ligada às variáveis Q4 “the media”, M4 “sailing, swimming, etc.” e H1 “architecture, houses and buildings”, mas trazendo uma crítica à ostentação e à aparência (*Screw house to wrecking flows, to BET on videos*). Essa mesma variável (I1) também está relacionada à variável S2 “people”, que mostra a ostentação para conseguir conquistar a outra pessoa (*Hot sex*

in back rows / I want to know what makes you scream / Be your twenty million dollar fantasy - Exemplo 1).

Exemplo 1 - Arquivo 77039 – Escore 57 – Título: Hollywood freaks – Artista: Beck (1999)

Letra da música	Variável
Champagne and ripple	B5
Shamans go cripple	
My sales go triple	I2
We drop lobotomy beats	
Evaporated meats	F1
On high-tech streets	Y2
We go solo	
Dance floors and talk shows	
Hot dogs , 'No Doz'	F1
Hot sex in back rows	S2
I want to know what makes you scream	
Be your twenty million dollar fantasy	I1
Treat you real good	
Expensive jeans	I1 / B5
Hollywood freaks on the Hollywood scene	Q4
Touch it real good if you want a piece	

A exaltação à riqueza também é vista na variável F2 “drinks and alcohol” (*Champagne and ripple* - Exemplo 1) e nas variáveis Y2 “information technology and computing” (*On high-tech streets* - Exemplo 1) e Y1 “science and technology in general”, também muito relacionadas com a variável B5, descrita acima (*Levis to Guess jeans, black and white to big screens* - Exemplo 3). O fato de querer ser melhor do que o outro se faz presente na variável I2 “business” (*My sales go triple* - Exemplo 1). Da mesma maneira, a variável L2 “living creatures: animals, birds, etc” se utiliza de animais para depreciação, exatamente como a variável M3 “vehicles and transport on land” (*From Bo-Bo's to Air Macs, low cars to horse backs* - Exemplo 3).

Exemplo 2 - Arquivo 176080 – Escore 55 – Título: Smokescreen – Artista: Citizen King (1999)

Letra da música	Variável
Now you eat the dust and you taste the puppy chow ain't no standstill hold	F1 / L2
Up captain crunch a desperado I take you out to lunch	F1
I'll yank your milk money	F2 / I1
Honey it's a royal treatment no tracks out	F1 / M3

Back in the wet cement a broke tooth And a phone booth and a change of clothes 'cause everyone knows I get Mean in the smoke screen half time going On the crime scene operation down low You call your hero but like sweetback I got no Afro I'm coming dark creepin' in	Q1 / H1 / B5 Q1 Q4
--	--------------------------

Já a variável Q1 “linguistic actions, states and processes, communication” mostra uma crítica ao comportamento das pessoas (*Bezel sets and baguettes, while talking on my cell phone* - Exemplo 3).

Exemplo 3 - 159655 – Escore 55 – Título: Remember when – Artista: H.S.E. (1999)

Letra da música	Variável
Bezel sets and baguettes , while talking on my cell phone From Bo-Bo's to Air Macs, low cars to horse backs Hooties to Cadillacs, water guns to chrome gats Levis to Guess jeans , black and white to big screens Six by nine to 18's, Bumsalack to sipping lean Regals with white walls, 20 inch with screens fall Michael Jackson off the wall, to MJG and 8Ball Screw house to wrecking flows , to BET on videos Zig-zags and yellow bags , to Glad sacks and optimos	F1 / Q1 M3 / L2 B5 / Y1 F2 Y1 H1 / M4 / I1 / Q4 B5

As variáveis que carregaram com mais peso estão relacionadas à comida, roupa, bebida e dinheiro, ou seja, bens que o dinheiro pode comprar e, pelos exemplos, vemos que há uma exaltação a esses objetos (Exemplo 3), depreciando uma outra pessoa a quem a música se refere (Exemplo 2).

Apesar de não ter o polo negativo, fomos observar exemplos do que não representa essa dimensão:

Exemplo 4 - Arquivo 67 – Escore 0 – Título: (“Till) I kissed you – Artista: The Everly Brothers (1960)

Never felt like this until I kissed you. How did I exist until I kissed you Never had you on my mind Now you're there all the time Never knew what I missed until I kissed you, uh-huh I kissed you, oh yeah Things have really changed since I kissed you, uh-huh
--

My life's not the same now that I kissed you, oh yeah
Mmm, you got a way about you
Now I can't live without you

Os Exemplos 4 e 5 mostram a fixação que o artista guarda pela pessoa amada (*How did I exist until I kissed you* - Exemplo 4) e não faz nenhuma crítica ou depreciação, nem pretende ser melhor do que ninguém. Também não há um anseio por bens materiais ou produtos de marca para se divertir (*The fun it is to be with you* - Exemplo 5), como ocorre nos Exemplos de 1 a 3, que são bem representativos do discurso da ostentação.

Exemplo 5 - Arquivo 73 – Escore 0 – Título: (All of a sudden) my heart sing – Artista: Bing Crosby (1914)

All of a sudden my heart sings
When I remember little things
The way you dance and hold me tight
The way you kiss and say goodnight
The crazy things we say and do
The fun it is to be with you
The magic thrill that's in your touch
Oh darling, I love so much!
The secret way you press my hand
To let me know you understand
The wind and rain on your face
The breathless world of your embrace

O Exemplo 6 já apresenta uma preocupação com o que está por vir, sem colocar o uso da tecnologia para ostentação nem querer ter o controle, mas uma aceitação do que está por vir (*Accept the unknown*).

Exemplo 6 - Arquivo 79 – Escore 0 – Título: (Another) 1984 – Artista: Billy Squier (1984)

From over the air comes a voice without care
Says we're doin' all right
But day out and day in from what he's been sayin'
We're in for a fight
I glance at myself and the world rushin' by
I can see for myself and the fear comes alive
Shades of termination—we're dyin' to survive
1984
Machine the mind it will rule by design
So you're never alone
We store information without confirmation
Accept the unknown

A presente dimensão, ao exaltar o discurso do materialismo, o faz de uma maneira descritiva e argumentativa; os textos que carregaram na dimensão tendem a conter um número maior de palavras para poder caracterizar essa ostentação marcadamente presente neles.

Os exemplos do extremo oposto do polo positivo mostram uma preocupação com o outro, sem a comparação do autor com a outra pessoa de forma depreciativa e sem o apelo aos bens materiais. Podemos dizer que há um contínuo, ilustrado na Figura 20:

FIGURA 20 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 1

57 - Ostentação de carros, comida e roupas
|
|
|
0 - Preocupação com o outro

fonte: elaborada pela autora

Então, nomeamos essa dimensão como:

Dimensão Semântica 1: Discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero.

4.1.2 - Fator 2, análise semântica

O Fator 2 possui três variáveis. Como a maioria dos estudos de análise fatorial, o segundo fator abarca menos variáveis que o primeiro, mas explica a segunda maior quantidade de variação do corpus, 2,51% que pode parecer pouco à primeira vista, porém as três variáveis estão intrinsecamente correlacionadas e aparecem juntas na amostra de textos envolvidos na inspeção feita na análise qualitativa. A seguir explicaremos cada uma das funções dessas variáveis, com exemplos do *corpus* para facilitar a compreensão.

A primeira variável, Q3 “language, speech and grammar”, mostra a presença de verbos como “read” (ler), no sentido de compreender a pessoa com quem o artista está falando (*Trying to read your mind is like driving when you're blind* - Exemplo 7), assim como a variável X1 “psychological actions, states and processes”, com substantivos como “mind” (mente), representando a pessoa com quem o artista está falando. A última variável que carregou nessa dimensão, X8 “trying” (tentativa), também se conecta nessa tentativa de entender a pessoa com quem o artista está falando.

Exemplo 7 - Arquivo 91297 – Escore 15 – Título: I’m trying to – Artista: Ashley Monroe (2018)

Letra da música	Variável
Trying to read your mind is like driving when you're blind	X8 // Q3 / X1
If only all this trying made it so	X8
How long do you try before you let it die?	X8
What do you need to know before you know?	
Trying to read your mind is like driving when you're blind	X8 // Q3 / X1
If only all this trying made it so	X8
I'm trying to God knows I am	X8
Make it different but it isn't	
I don't blame you for trying to change the truth	X8
I'm trying too.	X8
I know I should let you go but I can't leave you alone	
I'm trying to	X8

Já no Exemplo 8, a variável Q3 aparece como se fosse alguém que vai de encontro à pessoa com quem o artista está “conversando” (*Words might never come your way*).

Exemplo 8 - Arquivo 61603 – Escore 15 – Título: Gee and Ai – Artista: Bodyjar (1996)

Letra da música	Variável
Words might never come your way Because everything you do for me I would never do for you I can't tell you why 'Cause if I tried , I'd lie to you My one track mind , you're too kind Can't think of feelings I should show For what reason I don't know These things are hard to say You know they'll end up complicated Because everything you do for me I would never do for you. I can feel you wasting my time I can see you trying to read my mind	Q3 X8 X1 X8 / Q3 / X1

O Exemplo 9 traz as mesmas variáveis, mas há um não querer compreender nem ser compreendido pelas pessoas (*Can't hear a **word** they're sayin' / Only the echoes of my **mind***), há um querer se distanciar dos outros.

Exemplo 9 - Arquivo 51040 – Escore 15 – Título: Everybody's talkin' – Artista: Harry Nilsson (1968)

Letra da música	Variável
Can't hear a word they're sayin' Only the echoes of my mind People stoppin', starin' I can't see their faces Only the shadows of their eyes Goin' where the sun keeps shinin' In the pouring rain Goin' where the weather suits my clothes Banking off a northeast wind Sailin' on a summer's breeze Skipping over the ocean	Q3 X1 X8 X8

A partir da análise qualitativa cuidadosa de textos com pontuação notável nessa dimensão, notamos que há uma tentativa do artista de entender a outra pessoa a quem ele se dirige. O que os textos dessa dimensão possuem em comum é um sofrimento em decorrência da tentativa frustrada em entender o outro, além de uma descrição de processos mentais, como tentando explicar o que se passa no íntimo do artista (*Only the echoes of my **mind*** - Exemplo 9).

Apesar de não ter o polo negativo, fomos observar exemplos do que não representa essa dimensão:

Exemplo 10 - Arquivo 1 – Escore 0 – Título: #1 Crush – Artista: Garbage (1995)

I would die for you
I would die for you
I've been dying just to feel you by my side
To know that you're mine
I will cry for you
I will cry for you
I will wash away your pain with all my tears
And drown your fear
I will pray for you
I will pray for you
I will sell my soul for something pure and true
Someone like you
See your face ever place that I walk in

Nenhuma das variáveis carregou nos exemplos com escore 0 (já que não temos polo negativo, o escore 0 é o que podemos chamar de contraponto do polo positivo). Nenhum dos exemplos trata do drama interno nem de conflitos por falta de comunicação. Pelo contrário, o artista deixa claro que ele e seu interlocutor são bem conhecidos (*You know me, I don't need no introduction and shit* - Exemplo 11).

Exemplo 11 - Arquivo 4 – Escore 0 – Título: #1 Stunna – Artista: Juvenile Big Tymers Lil Wayne (2000)

Nigga can't out-stunt me when it come to these fuckin' cars, nigga
Believe that!
You know me, I don't need no introduction and shit
Ride Bentley's 'round the city on buttons, ya bitch
Arm hangin', wrist blingin' just stun'n and shit.
Drop the top block is hot
Stay bumpin', ya bitch
B. Atrice get it right, don't tangle and twist it
Hit the club every night drunk drinkin' that Crissy
Niggas mad don't like it 'cause I'm bangin' they bitches
When the light hit the ice, it twinkle and glistens
Baby, Brian B., Bubble you can call me what you feel

O mesmo ocorre no Exemplo 12, não há nenhum problema de comunicação (*May I remind you*) e que a pessoa a quem o artista se refere é conhecida de todos (*Everyone's seen the tricks you get up to*).

Exemplo 12 - Arquivo 8 – Escore 0 – Título: #17 – Artista: Hoodoo Gurus (2004)

May I remind you
May I remind you that you are through
Being the fresh blood
Everyone's seen the tricks you get up to
It's so nice, it's so cold
Your #17
You have a taste for something
You liked to very long ago
Nothing can equal blazing a trail
Through fields of virgin snow
It's so warm, it's so sweet
Your #17
It's so firm, it's so neat

A Dimensão Semântica 2, ao tratar do drama pessoal, conta uma história com muitos elementos narrativos. Os exemplos do extremo oposto do polo positivo (na verdade, os exemplos com escore 0, já que não há polo negativo) mostram que o autor conhece bem a pessoa com que ele está “falando” e sabe do que ele é capaz de fazer. Não há uma preocupação em querer entender o outro. Podemos dizer que há um contínuo:

FIGURA 21 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 2

15 - Procurar entender o outro e ser entendido pelas pessoas
|
|
|
0 - Conhecimento que o cantor tem do outro e que o outro possui dele

Fonte: elaborada pela autora

Então, nomeamos essa dimensão como:

Dimensão Semântica 2 – 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, descrição de processos mentais internos.

4.1.3 - Fator 3, análise semântica

O Fator 3 possui cinco variáveis. Como a maioria dos estudos de análise fatorial, o terceiro fator abarca menos variação que o primeiro e o segundo fatores, mas possui a terceira maior porcentagem de variação do *corpus*, 2,08%, que pode parecer pouco à primeira vista, porém as cinco variáveis estão intrinsecamente correlacionadas e aparecem juntas na amostra de todos os textos envolvidos na inspeção feita na análise qualitativa. A seguir, explicaremos cada uma das funções dessas variáveis, com exemplos do *corpus* para facilitar a compreensão.

A variável W3 “geographical terms” apresenta sentido figurado, onde termos geográficos, como “rivers” (rios) são utilizados de maneira metafórica (*When all the rivers turn to sand* - Exemplo 13). Podemos dizer que esse recurso é muito comum em músicas para trazer o discurso da poesia, do figurado para as letras. O mesmo ocorre com a variável W1 “universe” (*When there’s no sky above our land* - Exemplo 13).

Exemplo 13 - Arquivo 88324 – Escore 24 – Título: I’ll be over you – Artista: Danny (2007)

Letra da música	Variável
When all the rivers turn to sand	W3
When there’s no sky above our land	W1 / W3
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	W1 / W2
When angels no longer fly	S9 / M5
When tears run dry.	
I’ll be over you.	
When all the rivers turn to sand	W3
When there’s no sky above our land	W1 / W3

Já a variável W2 “light” aparece em seu sentido literal de iluminar, mas relacionada a outras variáveis que possuem sentido figurado (*When there’s no sun to light the day* - Exemplo 13, *Rain, sea, surf, sand, clouds then sun will shine* - Exemplo 14).

Exemplo 14 - Arquivo 144855 – Escore 23 – Título: Once more – Artista: Barclay James Harvest (1971)

Letra da música	Variável
Rain, sea, surf , sand, clouds then sun will shine Born in the spirit of love You took flight Soaring so high and so free No one can stop you You'll always be heard Our hearts know the truth In pools of blue I found you Let my love surround you Gave you life. From that day forever. Rain, sea, surf , sand, clouds then sun will shine Born in the spirit of love	W3 / W1 / W2 S9 M5 W3 W3 / W1 / W2 S9

A variável M5 “flying and aircraft” possui sentido literal de voar, assim como a variável S9 “religion and supernatural” (*When **angels** no longer **fly*** - Exemplo 13), que também possui o sentido de sagrado, abençoado (*Searching far away to the **holy land*** - Exemplo 15).

Exemplo 15 - Arquivo 198110 – Score 23 – Título: Riding on an arrow - Artista: Axel Rudi Pell (2008)

Letra da música	Variável
Searching far away to the holy land The closer we got without hesitation The sooner we found out the world is a land Time is passing on as long we ever here We won't stay long in the garden of fear Riding on an arrow, flying through the sky Riding on an arrow, flying on so high We're riding through the sky We fly in circle, trusting no one The shadows of the world are fading away Life's in the darkness kept on shining	S9 / W3 W1 / W3 M5 / W1 M5 W1 M5 W1 W2

Pelas variáveis que carregaram neste fator e pelos exemplos, podemos ver que tanto a natureza como a religião e o sobrenatural estão sendo usados de uma maneira figurada.

Apesar de não ter o polo negativo, fomos observar exemplos do que não representa essa dimensão:

Exemplo 16 - Arquivo 8 – Score 0 – Título: #17 – Artista: Hoodoo Gurus (2004)

May I remind you
May I remind you that you are through
Being the fresh blood
Everyone's seen the tricks you get up to
It's so nice, it's so cold
Your #17
You have a taste for something
You liked to very long ago
Nothing can equal blazing a trail
Through fields of virgin snow
It's so warm, it's so sweet
Your #17
It's so firm, it's so neat

Nenhuma das variáveis carregou nos exemplos com escore 0 (já que não temos polo negativo, o escore 0 é o que podemos chamar de contraponto do polo positivo). Nenhum dos exemplos trata do universo, nem do sobrenatural e, apesar de aparecer elementos da natureza de maneira figurada (*Through fields of virgin snow* - Exemplo 16), podemos dizer que o contexto onde este texto aparece é muito mais concreto do que figurado.

Algo curioso acontece com este texto do Exemplo 16, pois ele é o mesmo que o Exemplo 12, ou seja, ele é o oposto do que representa o discurso da tentativa de entender o outro e, ao mesmo tempo, o oposto do discurso do sobrenatural, ou seja, em textos como esse, pode-se ver uma concretude no que está sendo expresso sem uma preocupação com o outro.

Exemplo 17 - Arquivo 30 – Escore 0 – Título: Heroes – Artista: David Bowie (1977)

I will be king
And you, you will be queen
Though nothing will drive them away
We can beat them, just for one day
We can be heroes, just for one day
And you,
You can be mean.
And I, I'll drink all the time
'Cause we're lovers, and that is a fact
Yes we're lovers, and that is that
Though nothing, will keep us together

No Exemplo 17 pode-se ver que não há referência ao sobrenatural, nem ao místico e sim uma suposição do que pode acontecer (*I will be king / And you, you will be*

queen / Though nothing will drive them away / We can beat them, just for one day / We can be heroes, just for one day), enquanto que o Exemplo 18 traz uma concretude de maneira figurada, mas sem os elementos do universo (*False faces, bloody wallet / Ripping out your head, put on my clothes*).

Exemplo 18 - Arquivo 38 – Escore 0 – Título: Rollins Power Sauce – Artista: Beck (1994)

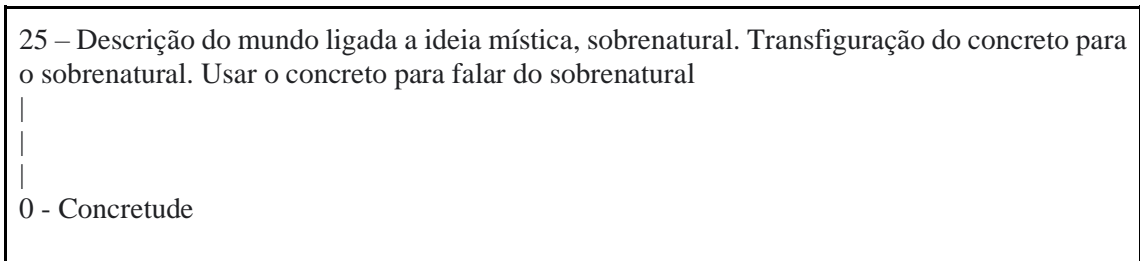
Umm, my throat muscles
Are completely shredded
You tell no one, nobody
Beating on you, no defenses
False faces, bloody wallet
Ripping out your head, put on my clothes
No, I wouldn't want you
Where the blood is all washed off
Where the blood is all washed off
And I got no focus, got no focus
Owhohwih no

A descrição do sobrenatural, assim como ocorre na Dimensão Semântica 1, que descreve o materialismo, é realizada também de maneira argumentativa, porém de uma forma mais concisa. Os textos possuem menos palavras do que os textos que carregaram na Dimensão Semântica 1.

Em uma dimensão, textos que se complementam são alocados juntos e isso fica claro na Dimensão Semântica 3, onde o aspecto figurado é muito forte nos seis exemplos citados aqui (Exemplos 13 a 18). A diferença fundamental é que os três primeiros (13, 14 e 15) são os mais representativos da dimensão – assim, além do aspecto figurado, todos os outros elementos estão presentes –, contrapondo-se aos demais exemplos, que possuem escore 0 e não apresentam essa característica, mas que guardam o seu sentido figurado, mostrando o pertencimento deles à dimensão.

Portanto, podemos dizer que há um contínuo:

FIGURA 22- DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 3



Fonte: elaborada pela autora

Como resultado, nomeamos essa dimensão como:

Dimensão Semântica 3 – Discurso sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente.

4.1.4 - Fator 4, análise semântica

O Fator 4 possui duas variáveis. Como a maioria dos estudos de análise fatorial, o quarto fator abarca menos variação que os três primeiros fatores, mas possui a quarta maior porcentagem de variação do *corpus*, 1,53%. As duas variáveis presentes no fator estão intrinsecamente correlacionadas e aparecem juntas na amostra de todos os textos envolvidos na inspeção feita na análise qualitativa. A seguir explicaremos cada uma das funções dessas variáveis, com exemplos do *corpus* para facilitar a compreensão.

Exemplo 19 - Arquivo 99952 – Score 10 – Título: It’s a whale – Artista: Deer Tick (2017)

Letra da música	Variável
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.	H4 M1 M1
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	H4 M1

Apesar de não ter o polo negativo, fomos observar exemplos do que não representa essa dimensão:

Exemplo 22 - Arquivo 38 – Escore 0 – Título: Rollins Power Sauce – Artista: Beck (1994)

Umm, my throat muscles
Are completely shredded
You tell no one, nobody
Beating on you, no defenses
False faces, bloody wallet
Ripping out your head, put on my clothes
No, I wouldn't want you
Where the blood is all washed off
Where the blood is all washed off
And I got no focus, got no focus
Owhohwih no.

O texto do Exemplo 22 é o mesmo texto Exemplo 18, ou seja, a concretude muito forte que aparece aqui e o afasta do discurso poético e sobrenatural (Dimensão Semântica 3), ao mesmo tempo em que o afasta do discurso do arrependimento e da saudade de um passado que não existe mais (*No, I wouldn't want you*).

Exemplo 23 - Arquivo 41 – Escore 0 – Título: The Take Over, the Breaks Over – Artista: Fall out boy (2009)

Baby, seasons change but people don't
And I'll always be waiting in the back room
I'm boring but overcompensate with
Headlines and flash, flash, flash photography
But don't pretend you ever forget about me
Don't pretend you ever forget about me
Wouldn't you rather be a widow than a divorcee?
Style your wake for fashion magazines
Widow or a divorcee?
Don't pretend it, don't pretend
We don't fight fair

No Exemplo 23, também fica claro que não há um arrependimento, e sim uma crítica ao comportamento da pessoa a quem o autor se dirige (*Don't pretend you ever forget about me / Wouldn't you rather be a widow than a divorcee?*), assim como aparece explicitamente no Exemplo 24 (*You think that you're so cool / That you're nobody's fool / But you've got a personality / Just like a bucket full of pee*).

Exemplo 24 - Arquivo 46 – Escore 0 – Título: ***hole – Artista: Gene Simmons (2004)

You think that you're so cool
That you're nobody's fool
But you've got a personality
Just like a bucket full of pee
How does it feel
To be a real asshole?
You're such a creep
You look like a sheep asshole
You know you've got no shame
And you've got such a stupid name
And one day you'll finally shut your trap
'Cause you are the cream of the crap

A Dimensão Semântica 4, assim como a Dimensão Semântica 2, se vale de elementos narrativos, pois o discurso do trágico-romântico é formado a partir da narração de uma história e possui elementos descritivos (como as dimensões 1 e 3) para tratar da busca da felicidade e do amor. Conforme relatamos, na Dimensão Semântica 3, as dimensões agrupam textos que se complementam. Nesta Dimensão Semântica 4, conseguimos ver um continuum que vai da crítica ao outro (textos como o Exemplo 23 e 24, que possuem escore 0) até uma saudade do que existiu, o discurso da saudade (Exemplo 21). Podemos então dizer que há um contínuo na dimensão, conforme mostra a Figura 23:

FIGURA 23 - DISCURSOS MAIS SALIENTES NA DIMENSÃO SEMÂNTICA 4

10 – Arrependimento e saudade de um passado que não existe mais
|
|
|
0 - Crítica ao outro

Fonte: elaborada pela autora

Então, nomeamos essa dimensão como:

Dimensão Semântica 4 – 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, da busca incessante pela felicidade.

Em suma, as letras de música apresentam quatro discursos bem distintos, que podem ser nomeados como:

- Dimensão Semântica 1 – Discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero;
- Dimensão Semântica 2 – 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, descrição de processos mentais internos;
- Dimensão Semântica 3 – Discurso sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente;
- Dimensão Semântica 4 – 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, da busca incessante pela felicidade.

As dimensões foram rotuladas com descritores longos e complexos, visto que as funções apresentadas pelas classes semânticas realizam discursos diferentes, com muitas facetas complementares. Sendo assim, a fim de condensar esses rótulos, propomos os seguintes rótulos sucintos, que capturam a essência discursiva de cada dimensão:

- Dimensão Semântica 1 - Dimensão do **materialismo**;
- Dimensão Semântica 2 - Dimensão do **existencialismo**;
- Dimensão Semântica 3 - Dimensão do **sobrenatural**;
- Dimensão Semântica 4 - Dimensão do **trágico-romântico**.

Podemos ver algumas complementaridades nestas dimensões. A Dimensão Semântica 1, ao exaltar o discurso do materialismo, o faz de uma maneira descritiva e argumentativa, da mesma maneira que a Dimensão Semântica 3, que descreve o sobrenatural. A Dimensão Semântica 2, ao tratar do drama pessoal, conta uma história com muitos elementos narrativos, assim como a Dimensão Semântica 4 se vale de elementos narrativos, mas também descritivos (como as dimensões 1 e 3), para tratar do trágico-romântico.

Ao inspecionarmos a correlação entre *tokens* (total de palavras) das letras de música e as dimensões, a Figura 21 nos mostra que há uma correlação moderada entre a Dimensão Semântica 1 e a quantidade de palavras na letra. Ou seja, quanto maior o número de palavras em uma letra de música, maior a tendência de ela estar fortemente representada na Dimensão Semântica 1 – o que faz sentido, pois se olharmos os textos que mais carregaram nessa dimensão, representados pelos Exemplos 1, 2 e 3 na Seção 4.1.1, percebe-se que eles contam uma história e que há uma descrição. Quase todos os textos que carregaram nessa dimensão pertencem aos gêneros hip hop ou rap, que necessitam de maior quantidade de palavras para retratar esse apelo à descrição do materialismo.

TABELA 7 – CORRELAÇÃO ENTRE AS DIMENSÕES E O TOTAL DE PALAVRAS DOS TEXTOS DO *CORPUS*

Coefficientes de Correlação de Pearson, N=201.065	
Prob > r sob H0: Rho=0	
	Total de palavras
Fator 1	0,4293 <,0001
Fator 2	0,11199 <,0001
Fator 3	-0,0742 <,0001
Fator 4	0,02522 <,0001

Fonte: elaborada pela autora

No extremo oposto da relação ilustrada na Tabela 7, vemos que o Fator 3 apresenta uma correlação negativa com o total de palavras nas letras das músicas, o que é curioso, visto que essa é uma dimensão descritiva, que retrata o sobrenatural, mas essa descrição ocorre de maneira mais concisa, com textos (letras) com uma quantidade de palavras bem menor do que as músicas da Dimensão Semântica 1.

A partir dessas quatro dimensões, podemos falar que a música em língua inglesa, em maior ou menor medida, aborda quatro discursos distintos, cuja caracterização sintética remete a representações diversas em torno do materialismo, do drama pessoal, do sobrenatural e do trágico-romântico.

4.2 RESULTADOS DA AMD ACÚSTICA

A partir dos resultados da análise fatorial, retornamos à Tabela 4, utilizada na metodologia e reproduzida novamente, na sequência. Tal tabela foi organizada em ordem decrescente de pesos das variáveis que carregaram em cada fator. Para melhor entendimento e facilitação da interpretação, apresentaremos a função que cada variável apresentou nos textos inspecionados na análise qualitativa. Ao contrário da Análise Multidimensional Semântica apresentada anteriormente neste capítulo, a Análise Multidimensional Acústica aqui apresentada inclui tanto polos positivos quanto negativos. Tal oposição, que na verdade vem a ser uma complementaridade, ou seja, em músicas onde as variáveis do polo positivo estão presentes, há uma grande chance de as variáveis do polo negativo não estarem presentes, e vice-versa. Os sinais + e - não possuem carga de valor, ou seja, um polo não é melhor que o outro, eles apenas indicam dados que se complementam, assim como os polos positivo e negativo de uma pilha.

TABELA 4 - PADRÃO FATORIAL DA ANÁLISE ACÚSTICA

Fator	Variável	Peso
F1+	energia	0,85
	volume	0,79
	batida por minuto	0,25
	vivacidade ²⁹	0,16
F1-	acusticidade	-0,74
F2+	dançabilidade	0,69
	valência	0,65
	compasso	0,24
	verbosidade	0,23
F2-	instrumentalidade	-0,28
F3+	modalidade	0,38
F3-	tom	-0,29

Fonte: elaborada pela autora

²⁹ Mostra a probabilidade de a faixa ter sido gravada em um show ao vivo.

4.2.1 - Fator 1, análise acústica

O Fator 1 possui cinco variáveis. Como a maioria dos estudos de análise fatorial, o primeiro fator abarca a maior porcentagem de variação do *corpus*, de 2,71% – relevante, embora não seja um valor alto. As cinco variáveis estão intrinsecamente correlacionadas e aparecem juntas na amostra de músicas envolvidas na inspeção feita na análise qualitativa. A seguir, explicaremos cada uma das funções dessas variáveis, com exemplos do *corpus*. Como não é possível descrever as características acústicas textualmente, indicamos o link da música na plataforma Spotify, para que o leitor possa acessá-la e verificar a presença das características acústicas mencionadas. As Tabelas 8, 9 e 10 descrevem as variáveis presentes no polo positivo do fator, enquanto a Tabela 11 descreve a variável que carregou no polo negativo.

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “energia” mostrou que a musicalidade marcada por essa variável pode ser entendida como intensidade, música alta, rápida, barulhenta, uma representação de atividade, com instrumentalidade amplificada. A Tabela 8, a seguir, apresenta exemplos de canções com alto valor de energia. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 8 - CANÇÕES COM ALTO VALOR DE ENERGIA

Título	Artista	Valor	Link do Spotify
White Riot – Live	Cock Sparrer	0,99	https://open.spotify.com/track/34iEDajZdKv_oDBxPnaen1y
Self vs. Self	Pendulum in Flames	0,99	https://open.spotify.com/track/2lN6G35gsXkA3xzPYqmis5
The American in me (live)	The Avengers	0,94	https://open.spotify.com/track/1R0PdvKuCp_gAA5GFfK6pKY

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “batidas por minuto” mostrou que a musicalidade marcada por essa variável pode

ser entendida como velocidade, o que dá o ritmo da música. A Tabela 9, a seguir, apresenta exemplos de canções com alto valor de batidas por minuto. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 9 - CANÇÕES COM ALTO VALOR DE BATIDAS POR MINUTO

Título	Artista	Valor	Link do Spotify
White Riot – Live	Cock Sparrer	217,54	https://open.spotify.com/track/34iEDajZdKvoDBxPnaenIy
The American in me (live)	The Avengers	201,78	https://open.spotify.com/track/1R0PdvKuCpgAA5GFfK6pKY

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “volume” mostrou que a musicalidade marcada por essa variável pode ser entendida como volume alto, referente à força da música. Essa variável aparece com valores negativos. Portanto, quanto mais próximo de 0, mais alta a música é. A Tabela 10, a seguir, apresenta um exemplo de canção com alto valor de volume. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 10 - CANÇÃO COM ALTO VALOR DE VOLUME

ítulo	Artista	Valor	Link do Spotify
Self vs. Self	Pendulum in Flames	-2,35	https://open.spotify.com/track/2IN6G35gsXkA3xzPYqmis5

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “vivacidade” mostrou que a musicalidade marcada por essa variável pode ser entendida como interação com o público, mostrando que a faixa foi gravada ao vivo. A Tabela 11, a seguir, apresenta exemplos de canções com alto valor de vivacidade. A tabela

mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 11 - CANÇÕES COM ALTO VALOR DE VIVACIDADE

Título	Artista	Valor	Link do Spotify
White Riot – Live	Cock Sparrer	0,96	https://open.spotify.com/track/34iEDajZdKv_oDBxPnaen1y
Self vs. Self	Pendulum in Flames	0,94	https://open.spotify.com/track/2IN6G35gsXkA3xzPYqmis5
The American in me (live)	The Avengers	0,93	https://open.spotify.com/track/1R0PdvKuCp_gAA5GFFK6pKY

Fonte: elaborada pela autora

Agora passaremos à análise da única variável que carregou no pólo negativo. A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “acusticidade” mostrou que a musicalidade marcada por essa variável pode ser entendida como música totalmente acústica (valor 1), sem letras, o que não é o caso de nenhuma música do *corpus* de estudo, mas algumas faixas ficam bem próximas desse valor. As músicas com valores altos de acusticidade tendem a ser mais introvertidas e contemplativas, com sensibilidade vocal, suavidade, calma e sutileza de expressão. A Tabela 12, a seguir, apresenta exemplos de canções com alto valor de acusticidade. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 12 - CANÇÕES COM ALTO VALOR DE ACUSTICIDADE (INSTRUMENTAÇÃO)

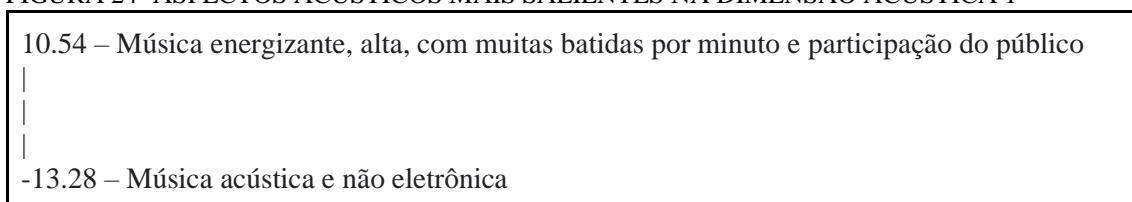
Título	Artista	Valor	Link do Spotify
Outside the Wall – Remastered Version	Pink Floyd	0,62	https://open.spotify.com/track/1zeyA74WeKOpYH6wtXQwkp
It’s all so new to me	Jan Savitt	0,99	https://open.spotify.com/track/26G0R3rwgX0kE68QOABFvt

Translucence	Derek Jarman	0,97	https://open.spotify.com/track/1Kof3REYOriwF6WjJ7IHsu
--------------	--------------	------	---

Fonte: elaborada pela autora

As músicas estão em um contínuo:

FIGURA 24- ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 1



fonte: elaborado pela autora

Então nomeamos a dimensão como:

Dimensão Semântica 1 – Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo vs Musicalidade com maior ênfase na instrumentação e em ritmo ameno, mais introvertida e contemplativa, com sensibilidade vocal, suavidade, calma e sutileza de expressão.

4.2.2 - Fator 2, análise acústica

O Fator 2 também possui cinco variáveis. Como outros estudos de análise fatorial, o segundo fator abarca a segunda maior porcentagem (1,56%) de variação do *corpus*. As cinco variáveis estão intrinsecamente correlacionadas, e aparecem juntas na amostra de músicas envolvidas na inspeção feita na análise qualitativa. A seguir explicaremos cada uma das funções dessas variáveis, com exemplos do *corpus*. Como não é possível descrever as características acústicas textualmente, indicamos o link da música na plataforma Spotify, para que o leitor possa acessá-la e verificar a presença das características acústicas mencionadas. As Tabelas 13, 14, 15 e 16 descrevem as variáveis presentes no polo positivo do fator, enquanto a Tabela 17 descreve a variável que carregou no polo negativo.

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “compasso” mostrou que a musicalidade marcada por essa variável pode ser entendida como música convencional, no compasso do que chamamos música contemporânea. A Tabela 13, a seguir, apresenta exemplos de canções com alto valor de compasso. A tabela mostra o título da música, artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 13 - CANÇÕES COM ALTO VALOR DE COMPASSO

Título	Artista	Valor	Link do Spotify
I'm special (Skit) (Explicit)	Mack 10	5	https://open.spotify.com/track/7GdKNXWCtofuATCoFW38Z8
Wake up Mr. West	Kanye West	5	https://open.spotify.com/track/62KeM7jbmYvzT5FycdiQeS
Sister surprise	Gary Numan	5	https://open.spotify.com/track/6VY5SleF8ASIIW6OLzWBib

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “verbosidade” mostrou que a musicalidade marcada por essa variável pode ser muito próxima a um discurso ou a podcasts – onde não há música, mas apenas uma ou mais pessoas falando. A Tabela 14, a seguir, apresenta exemplos de canções com um alto valor de verbosidade. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 14 - CANÇÕES COM ALTO VALOR DE VERBOSIDADE

Título	Artista	Valor	Link do Spotify
I'm special (Skit) (Explicit)	Mack 10	0,91	https://open.spotify.com/track/7GdKNXWCtofuATCoFW38Z8
Wake up Mr. West	Kanye West	0,93	https://open.spotify.com/track/62KeM7jbmYvzT5FycdiQeS

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “valência” mostrou que a musicalidade marcada por essa variável pode ser entendida como música alegre, feliz e eufórica. A Tabela 15, a seguir, apresenta um exemplo de canção com alto valor de valência. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 15 - CANÇÃO COM ALTO VALOR DE VALÊNCIA

Título	Artista	Valor	Link do Spotify
Sister surprise	Gary Numan	0,90	https://open.spotify.com/track/6VY5SleF8ASIIW6OLzWB1b

Fonte: elaborada pela autora

A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “dançabilidade” mostrou que a musicalidade marcada por essa variável pode ser entendida como música para dançar, repetitiva, rítmica e com força. A Tabela 16, a seguir, apresenta exemplos de canções com alto valor de dançabilidade. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para audição da faixa.

TABELA 16 - CANÇÕES COM ALTO VALOR DE DANÇABILIDADE

Título	Artista	Valor	Link do Spotify
I'm special (Skit) (Explicit)	Mack 10	0,74	https://open.spotify.com/track/7GdKNXWCTofuATCoFW38Z8
Wake up Mr. West	Kanye West	0,75	https://open.spotify.com/track/62KeM7jbmYvzT5FycdiQeS
Sister surprise	Gary Numan	0,67	https://open.spotify.com/track/6VY5SleF8ASIIW6OLzWB1b

Fonte: elaborada pela autora

A seguir explicaremos as funções da única variável que carregou no polo negativo deste fator, com exemplos do *corpus*. A análise qualitativa de muitos exemplares

do *corpus* com valores notáveis da variável “instrumentalidade” mostrou que a musicalidade marcada por essa variável pode ser entendida como música instrumental, com pouco conteúdo vocal, maior sensibilidade, intimidade, profundidade e subjetividade. A Tabela 17, a seguir, apresenta exemplos de canções com alto valor de instrumentalidade. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para audição da faixa.

TABELA 17 - CANÇÕES COM ALTO VALOR DE INSTRUMENTALIDADE

Título	Artista	Valor	Link do Spotify
Sacrificed	Napalm Death	0,88	https://open.spotify.com/track/4zjJ9fVE6ZWH5pffeEaQdK
Time = Cause	Broken Social Scene	0,93	https://open.spotify.com/track/5VHaHZIofYj0VtF2eZmnhl
The story of a soldier	Ennio Morricone	0,89	https://open.spotify.com/track/3cwREafAVNjUEt0hdrNVkt

Fonte: elaborada pela autora

As músicas estão num contínuo:

FIGURA 25 - ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 2

<p>15,67 – Música rítmica, repetitiva, alegre, convencional, com conteúdo vocal, feita para dançar</p> <p> </p> <p> </p> <p> </p> <p>-9,31 – Música instrumental, com pouco conteúdo vocal</p>
--

Fonte: elaborada pela autora

Então nomeamos a dimensão como:

Dimensão Acústica 2 – Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade versus Musicalidade voltada para a menor predominância vocal, maior sensibilidade, intimidade, profundidade, subjetividade.

4.2.3 - Fator 3, análise acústica

O Fator 3 também possui duas variáveis. O terceiro fator abarca a terceira maior porcentagem de variação do *corpus*, 1,20%. As duas variáveis que o compõem se complementam: uma aparece no pólo positivo e a outra, no negativo. Como não é possível descrever as características acústicas textualmente, indicamos o link da música na plataforma Spotify, para que o leitor possa acessá-la e verificar a presença das características acústicas mencionadas. A Tabela 18 traz exemplos da variável presente no polo positivo do fator, enquanto a Tabela 19 ilustra a variável que carregou no polo negativo.

A seguir, explicaremos as funções da única variável que carregou no polo positivo deste fator, com exemplos do *corpus*. A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “modalidade” mostrou que a musicalidade marcada por essa variável pode ser entendida como música menos melancólica, com acordes maiores, mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental. A Tabela 18 a seguir apresenta exemplos de canções com alto valor de modalidade. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

TABELA 18 - CANÇÕES COM VALOR MÁXIMO (1) DE MODALIDADE

Título	Artista	Valor	Link do Spotify
Hello Dolly!	Louis Armstrong	1	https://open.spotify.com/track/63kd4m3VFxcJjPVVtbVNAu
Spiritual Cramp [Symptom Reversal] (Remixed by Joseph Biashara)	Christian Death	1	https://open.spotify.com/track/7eJ6bkzHjG EsMioychFWry
Fireal	Deftones	1	https://open.spotify.com/track/7y0MUu7TzIA7ZyieRsNK4d

Fonte: elaborada pela autora

A seguir, explicaremos as funções da única variável que carregou no polo negativo deste fator, com exemplos do *corpus*. A análise qualitativa de muitos exemplares do *corpus* com valores notáveis da variável “tom” mostrou que a musicalidade marcada por essa variável pode ser entendida como música baseada em acordes mais tristes e tons mais baixos, com sonoridade mais emotiva e temperamental. A Tabela 19, a seguir,

apresenta exemplos de canções com alto valor de tom. A tabela mostra o título da música, o artista responsável pela gravação, o valor da medida acústica correspondente e o link do Spotify para a audição da faixa.

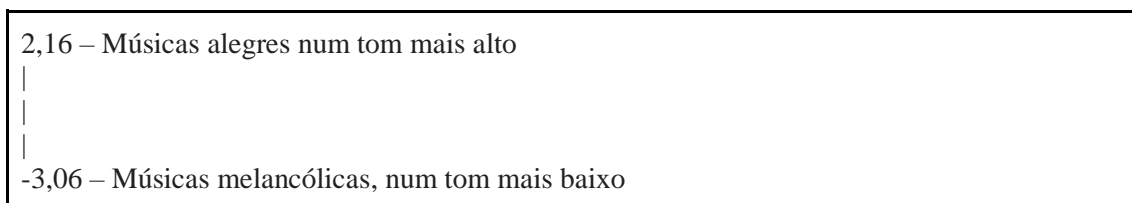
TABELA 19 - CANÇÕES COM ALTO VALOR DE TOM

Título	Artista	Valor	Link do Spotify
Give it a chance	Foxwarren	11	https://open.spotify.com/track/7nM1sBdMH8GHaZbQIIYQSI
I hate today	Godhead	11	https://open.spotify.com/track/3YXvwbDn5GOYwi4xncGxjh
Bus driver	Muddy Waters	11	https://open.spotify.com/track/6aMqOnAy0Q5EkHyjib4VXE

Fonte: elaborada pela autora

As músicas estão num contínuo:

FIGURA 26 - ASPECTOS ACÚSTICOS MAIS SALIENTES NA DIMENSÃO ACÚSTICA 3



Fonte: elaborada pela autora

Então nomeamos a dimensão como:

Dimensão Acústica 3 – Musicalidade baseada em acordes mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental versus Musicalidade baseada em acordes menos alegres e tons mais baixos, com sonoridade mais emotiva e temperamental

A musicalidade expressa pelas características acústicas analisadas pode ser representada por três dimensões acústicas, conforme mostramos acima. Essas três dimensões são:

- Dimensão Acústica 1 – Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo versus Musicalidade com maior ênfase na instrumentação e em ritmo ameno, mais introvertida e contemplativa, com sensibilidade vocal, suavidade, calma e sutileza de expressão.
- Dimensão Acústica 2 – Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade versus Musicalidade voltada para a menor predominância vocal, maior sensibilidade, intimidade, profundidade, subjetividade.
- Dimensão Acústica 3 – Musicalidade baseada em acordes mais alegres e tons mais baixos, com sonoridade menos emotiva e menos temperamental versus Musicalidade baseada em acordes mais tristes e tons mais altos, com sonoridade mais emotiva e temperamental.

As dimensões foram rotuladas com descritores longos e complexos, visto que as funções apresentadas pelas classes acústicas realizam musicalidades diferentes, com muitas facetas complementares. Sendo assim, a fim de condensar esses rótulos, propomos os seguintes rótulos sucintos, que capturam a essência musical de cada dimensão:

- Dimensão Acústica 1 - **Musicalidade intensa versus intimista;**
- Dimensão Acústica 2 - **Musicalidade dançante versus contemplativa;**
- Dimensão Acústica 3 - **Musicalidade alegre versus triste.**

A fim de detectar o grau de independência das dimensões acústicas, conduzimos uma análise de correlação, cujos resultados aparecem na Tabela 20, a seguir. Podemos notar que as Dimensões Acústicas 1 e 2 estão altamente relacionadas, ou seja, musicalidade intensa (polo positivo da Dimensão Acústica 1) relaciona-se ligeiramente

com musicalidade para dançar (polo positivo da Dimensão Acústica 2), enquanto músicas mais intimistas (polo negativo da Dimensão Acústica 1) estão relacionadas com músicas feitas para sentir (polo negativo da Dimensão Acústica 2).

TABELA 20 - CORRELAÇÃO ENTRE AS DIMENSÕES ACÚSTICAS

Correlações Inter-Fatoriais			
	Fator 1	Fator 2	Fator 3
Fator 1	1,00	0,30	-0,15
Fator 2	0,30	1,00	-0,09
Fator 3	-0,15	-0,09	1,00

Fonte: elaborada pela autora

Como as demais correlações são baixas, podemos concluir que as dimensões acústicas são em grande parte independentes, ou seja, cada uma representa uma forma musical em boa medida distinta das demais.

4.3 ANÁLISE DE CORRELAÇÃO CANÔNICA

A partir dos resultados da análise canônica, por meio da qual são calculadas correlações entre os escores das canções nas dimensões semânticas e acústicas, previamente explicadas (Seções 4.1 e 4.2), chegamos a três variantes canônicas (*canonical variates*), que foram interpretadas como três dimensões musicais. O resultado geral apresentado pelo procedimento CANCERR do SAS OnDemand for Academics mostra que as três correlações atingiram valor significativo de p , como mostra a Tabela 21, a seguir. No entanto, todas possuem correlações de valor mediano (a primeira), ou baixo (as duas outras). Do mesmo modo, as três correlações canônicas capturam pouca variação (9,3% para a primeira, 1,7% para a segunda e 0,1% para a terceira). Porém, ao inspecionar as dimensões correlacionadas, decidimos pela inclusão das três correlações canônicas, a despeito de não serem fortes, pois levamos em conta que as três correlações indicavam uma relação interpretável entre as dimensões semânticas e acústicas.

TABELA 21 - CORRELAÇÃO CANÔNICA ENTRE AS DIMENSÕES SEMÂNTICAS E ACÚSTICAS

	Correlação canônica	F aproximado	<i>p</i>	Correlação canônica ao quadrado	Porcentagem de variação capturada
1	0,30	835,63	<0,0001	0,0929	9,29%
2	0,04	30,52	<0,0001	0,0017	1,7%
3	0,01	6,05	0,0024	0,0001	0,1%

Fonte: elaborada pela autora

A Tabela 22 apresenta os pares de correlações canônicas entre as dimensões semânticas e acústicas.

TABELA 22 - ESTRUTURA CANÔNICA

Correlações entre as Dimensões Semânticas e suas Variáveis Canônicas			
	Variável Canônica Semântica 1	Variável Canônica Semântica 2	Variável Canônica Semântica 3
Dimensão Semântica 1	0,818	0,313	0,147
Dimensão Semântica 2	0,101	0,807	-0,235
Dimensão Semântica 3	-0,530	0,512	0,061
Dimensão Semântica 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: elaborada pela autora

Para obter os pares canônicos, temos que observar a coluna Variável Canônica Semântica 1 e Variável Canônica Acústica 1 (Tabela 22) e selecionar correlações maiores ou iguais a 0,30 (ponto de corte). Podemos observar que o primeiro par de variantes canônicas é formado pelas dimensões Dimensão Semântica 1 (0,81), Dimensão Semântica 3 (-0,53) e Dimensão Acústica 2 (0,99). Já o segundo par de variantes canônicas é formado pelas dimensões Dimensão Semântica 1 (0,31), Dimensão Semântica 2 (0,80), Dimensão Semântica (0,51), Dimensão Acústica 1 (0,91) e Dimensão

Acústica 3 (-0,35). E, finalmente, o terceiro par de variantes canônicas é formado pelas dimensões Dimensão Semântica 4 (0,96) e Dimensão Semântica 3 (0,93).

A seguir detalharemos cada um desses pares de variantes canônicos.

4.3.1 - Primeiro par de variantes canônicas

O primeiro par de variantes canônicas é formado pela Dimensão Semântica 1, pelo contrário do que a Dimensão Semântica 3 representa (já que ela não possui polo negativo) e pela Dimensão Acústica 2. A Tabela 23, na sequência traz cada uma dessas variáveis com a definição do que cada uma representa.

TABELA 23 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO PRIMEIRO PAR DE VARIANTES CANÔNICAS

Dimensões Semânticas	Dimensões Acústicas
Dimensão Semântica 1 – Discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero (0,81)	Dimensão Acústica 2 – Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade (0,99)
Dimensão Semântica 3 – Exceto o discurso sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente (-0,53)	

Fonte: elaborada pela autora

Como a Dimensão Acústica 2 é a variável com maior peso no primeiro par de variantes canônicas, essa é a que influencia mais a dimensão, com maior peso do que as dimensões semânticas.

A Dimensão Acústica 2 possui o maior peso nesse par de variantes canônicas. Conforme mostramos acima, a análise qualitativa de muitos exemplares do *corpus* com valores notáveis dessa variável mostrou que a musicalidade marcada por essa variável pode ser entendida como música baseada em ritmo repetitivo, voltada para expressão corporal em ambientes fechados, com predominância de vocalização e verbosidade. A seguir, apresentaremos exemplos de músicas que mais carregaram no par de variantes canônicas, com os respectivos links do Spotify para a audição das faixas.

O discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal e do efêmero pode ser visto nas músicas que carregaram nessa dimensão.

O Exemplo 43 mostra o quanto o artista não se importa com a sociedade. Essa música faz parte de um musical da Broadway, como uma paródia do seriado Vila Sésamo e trata de temas como racismo, homossexualidade e pornografia na internet (*I'm not wearing underwear today*).

Exemplo 43 – Arquivo 15219 – Título: I'm not wearing underwear today – Artista: Original Broadway Cast of Avenue Q (2003) <https://open.spotify.com/track/72oRqFFeYCML3FRBuvjcJm> - Escores: Dimensão Semântica 1 positiva: 10; Dimensão Semântica 3 negativa: 0; Dimensão Acústica 2 positiva: 14,13

Letra da música	Variável
I'm not wearing underwear today	Dim 1 Sem
No, I'm not wearing underwear today	Dim 1 Sem
Not that you prob'ly care much about my underwear	Dim 1 Sem
Still, nonetheless, I gotta say	
That I'm not wearing underwear to-daaaay!	Dim 1 Sem
Christmas eve	
Get a job!	
Thank you honey	Dim 1 Sem

Já o Exemplo 44 descreve um bairro com destaque à opulência e às pessoas que dão valor ao materialismo (*Kids with their spats / Ladies with chauffeurs / Dogs wearing hats and jackets / Rich apartments*). Este exemplo também mostra como o autor se identifica com esse materialismo (*I love the exquisite array / I love the camp as camp parade*).

Exemplo 44 – Arquivo 129752 – Título: Mornington crescent – Artista: Belle and Sebastian (2006) - Alternative/Indie: <https://open.spotify.com/track/2jHfc8NdjF0TLyIIXPV4Rn> - Escores: Dimensão Semântica 1 positiva: 42; Dimensão Semântica 3 negativa: 0; Dimensão Acústica 2 positiva: 3,19

Letra da música	Variável
-----------------	----------

Kids with their spats	Dim 1 Sem
Ladies with chauffeurs	Dim 1 Sem
Dogs wearing hats and jackets	Dim 1 Sem
Rich apartments	Dim 1 Sem
Old punk posters	Dim 1 Sem
Tartan garments	Dim 1 Sem
I love the exquisite array	
I love the camp as camp parade	
The possibilities suggest themselves to me	
I'm feeling free	
Mornington Crescent	
The sun in the east	

A Dimensão semântica 3 não carregou neste par de variantes canônicas. Nota-se a sua ausência, isto é, a falta de itens lexicais que remetem ao discurso do sobrenatural, do místico, do figurado, do lírico, do transcendente. Ao inspecionarmos os textos que carregaram neste par de variante canônico encontramos uma apelo muito forte ao concreto totalmente contrário ao lirismo que a Dimensão semântica 3 representa (*There's easy **money**, jobs / Especially when you **build** the bombs / That blow big cities off the **map** / Just guess who profits when we **build'em** back up* - Exemplo 45).

Exemplo 45 – Arquivo 224499 – Título: When ya get drafted – Artista: Dead Kennedys (1980) – Rock: <https://open.spotify.com/track/2LiywNrgaig5oCINlp6WcF> - Escores: Dimensão Semântica 1 positiva: 42; Dimensão Semântica 3 negativa: 0; Dimensão Acústica 2 positiva: 5,18

Letra da música	Variável
And Russians nibbling everywhere	Dim Sem 1
The chessboard's filling up with red	Dim Sem 1
We make more profits when we blow off their heads	
Economy is looking bad,	Dim Sem 1
Let's start another war, when ya get drafted	Dim Sem 1
Fan the fires of racist hatred, we want total war, when ya get drafted!	Dim Sem 1
Drooling fingers, panic buttons	Dim Sem 1
Playing with missiles like they're toys	
There's easy money , jobs	Dim Sem 1
Especially when you build the bombs	Dim Sem 1
That blow big cities off the map	Dim Sem 1
Just guess who profits when we build'em back up	Dim Sem 1

A partir das variáveis carregadas no par de variante canônico e de uma análise qualitativa cuidadosa baseada em muitos exemplares de texto e de áudio, nomeamos a primeira Dimensão canônica como:

Dimensão musical 1 – Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade, retratando o discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero distanciando-se do sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente.

4.3.2 - Segundo par de variantes canônicas

O segundo par de variantes canônicas é formado pela Dimensão Semântica 1, Dimensão Semântica 2, Dimensão Semântica 3, pela Dimensão Acústica 1 e pela Dimensão Acústica 3 negativa. A Tabela 24, abaixo, traz cada uma dessas variáveis com a definição do que cada uma representa.

TABELA 24 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO SEGUNDO PAR DE VARIANTES CANÔNICAS

Dimensões Semânticas	Dimensões Acústicas
Dimensão Semântica 1 positiva – Discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero (0,31)	Dimensão Acústica 1 positiva – Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo (0,91)
Dimensão Semântica 2 positiva – 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, descrição de processos mentais internos (0,80)	Dimensão Acústica 3 negativa – Musicalidade baseada em acordes mais tristes e tons mais altos, com sonoridade mais emotiva e temperamental (-0,35)
Dimensão Semântica 3 positiva – Discurso sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente (0,51)	

Fonte: elaborada pela autora

<https://open.spotify.com/track/3FLLxeumHAWKBgphK8R1Lx> - Escores: Dimensão Semântica 1 positiva: 10; Dimensão Semântica 2 positiva: 15; Dimensão Semântica 3 positiva: 5; Dimensão Acústica 1 positiva: 4,19; Dimensão Acústica 3 negativa: -0,08

Letra da música	Variável
Television eye hypnotizing me in my bed I've got no time to sleep, my eyes are getting blood-shot red You've got me in your power, staring at your many channeled head	Dim Sem 1
Television eye before you got me hooked I used to read Now you're by my bedside shining empty lights on me	Dim Sem 1 e 2 Dim Sem 3
'Round the clock commercials, trying to sell me things I'll never need It's hard to turn you off, even though you never turn me on	Dim Sem 1 e 2
You're trying to slow my thinking, feeling like my mind is nearly gone	Dim Sem 2
Television eye, had me in your power too long	Dim Sem 1

A partir da inspeção dos textos, verificou-se uma presença marcante do discurso da descrição de processos mentais internos (Dimensão Semântica 2), mas com duplo sentido, conforme mostra o Exemplo 48: enquanto para alguns há uma luta, um conflito interno para entender a si mesmo e ao outro (*Some **people struggle** daily / They **struggle** with their conscience*), ocorre também uma aceitação do que a pessoa é, sem nenhum conflito interno, pelo menos o autor quer que pareça que não há um conflito e que ele aceita sua vida sem nenhum problema existencial (*I have no guilt to **haunt** me / I feel no wrong intent*).

Exemplo 48 – Arquivo 2018 – Título: A criminal mind – Artista: Gowan (1985) – Alternative / Indie: <https://open.spotify.com/track/6cwpu9CaBztS9BO2z1m80z> - Escores: Dimensão Semântica 1 positiva: 10; Dimensão Semântica 2 positiva: 15; Dimensão Semântica 3 positiva: 2; Dimensão Acústica 1 positiva: 0,9; Dimensão Acústica 3 negativa: -1,94.

Letra da música	Variável
Before you read my sentence I'd like to say a few words Here in my own defense	Dim Sem 2 Dim Sem 2
Some people struggle daily They struggle with their conscience Till the end	Dim Sem 1 e 2 Dim Sem 2
I have no guilt to haunt me I feel no wrong intent I'm made of cold stone Just like your prison walls	Dim Sem 3

A criminal mind I AM	Dim Sem 2
--------------------------------	-----------

A partir das variáveis carregadas no par de variante canônico e de uma análise qualitativa cuidadosa baseada em muitos exemplares de texto e de áudio, nomeamos a segunda Dimensão Canônica como:

Dimensão Musical 2 – Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo, com 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, descrição de processos mentais internos, descrevendo o mundo como sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente.

4.3.3 - Terceiro par de variantes canônicas

O terceiro par de variantes canônicas é formado pela Dimensão Semântica 4 e pela Dimensão Acústica 3 positiva. A Tabela 25, na sequência traz cada uma dessas variáveis com a definição do que cada uma representa.

TABELA 25 - DIMENSÕES SEMÂNTICAS E ACÚSTICAS CORRELACIONADAS DO TERCEIRO PAR DE VARIANTES CANÔNICAS

Dimensão Semântica	Dimensão Acústica
Dimensão Semântica 4 – 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, da busca incessante pela felicidade (0,96)	Dimensão Acústica 3 positiva - Musicalidade baseada em acordes mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental (0,93)

fonte: elaborado pela autora

Como a Dimensão Semântica 4 é a variável com maior peso na Dimensão Canônica 3, ela é a que maior influência possui neste par de variantes canônicas, seguida da Dimensão Acústica 3. A seguir, apresentaremos exemplos de músicas que mais

carregaram no terceiro par de variantes canônicas.

A inspeção dos textos desse par de variantes canônicas revelou o discurso do trágico-romântico da jornada pessoal, do sofrimento pela solidão, retratado no Exemplo 49 (*My life is so empty, nothing to **live** for / My mind is all confusion, 'cos I defied the law / When you weren't there to help me, I lost my mind and **ran***).

Exemplo 49 – Arquivo 98116 – Título: Innocent Exile – Artista: Iron Maiden (1981) – Metal: <https://open.spotify.com/track/39uocZ5W2VV74lciqMR4lg> - Escores: Dimensão Semântica 4 positiva: 10; Dimensão Acústica 3 positiva: 2,16.

Letra da música	Variável
When you weren't there to help me, I lost my mind and ran	Dim Sem 4
I never had no trouble before this all began	
My life is so empty, nothing to live for	Dim Sem 4
My mind is all confusion, 'cos I defied the law	
When you weren't there to help me, I lost my mind and ran	Dim Sem 4
I never had no trouble before this all began	
I'm running away , nowhere to go	Dim Sem 4
I'm lost and tired and I just don't know	
Yeah	
They say I killed a woman, they know it isn't true	
They're just trying to frame me, and all because of you. Yeah	
Yeah	

Textos retratando o mito do lar ideal, onde não há problemas, só o lado bom da vida também carregaram neste par de variantes canônicas, como retrata o Exemplo 50 (*But luckily I, I got a mind to know / On my way back **home***).

Exemplo 50 – Arquivo 144209 – Título: On my way back home – Artista: Band of Horses (2010) - Alternative / Indie: <https://open.spotify.com/track/5qJw5V16ISycNIENMO1RD0> - Escores: Dimensão Semântica 4 positiva: 10; Dimensão Acústica 3 positiva: 2,16

Letra da música	Variável
The only words that could think of	
I'm pissing my life away in the form of a song	
On my way back home	Dim Sem 4
Every step , a victory it was	Dim Sem 4
I was cheating death, just in time I woke	
My memories start to wander off	Dim Sem 4
Come to me, the remembrance of	
On my way back home	Dim Sem 4

I came in this way and here now I'll stay	Dim Sem 4
If the unknown have to wait one more day	
There's often times that it comes out wrong	Dim Sem 4
But luckily I, I got a mind to know	
On my way back home	Dim Sem 4

Textos como o Exemplo 51 retratam o discurso do amor idealizado (*Imagination of you and I*) e do reencontro (*Try to be patient, here I come*) também são comuns em textos do terceiro par de variantes canônicos.

Exemplo 51 – Arquivo 3159 – Título: A season in green – Arquivo: Besech (2002) – Metal: <https://open.spotify.com/track/5ZyJFkpPPxhzJxl1OvI129> - Escores: Dimensão Semântica 4 positiva: 10; Dimensão Acústica 3 positiva: 2,16

Letra da música	Variável
And a thought came sealing down	Dim Sem 4
Imagination of you and I	
So much depression by losing you	
A new beginning without grief	
Painfully a season in green came	Dim Sem 4
And our home went up in grief	Dim Sem 4
Someone came to bring us the message	
Saying that there	
Is nothing to do	
Do you have name my child	
Maybe that day something died	
Where are you now cold in a cage	
Try to be patient, here I come	Dim Sem 4

A partir das variáveis carregadas no par de variante canônico e de uma análise qualitativa cuidadosa baseada em muitos exemplares de texto e de áudio, nomeamos a terceira dimensão canônica como:

Dimensão Musical 3 – Musicalidade baseada em acordes mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental retratando 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, da busca incessante pela felicidade.

Em suma, identificamos três dimensões musicais bem distintas, nomeadas como:

- Dimensão Musical 1 – Musicalidade baseada em ritmo repetitivo, vibrante, voltada para expressão corporal em ambientes fechados, em compasso convencional, com predominância de vocalização e verbosidade, retratando o discurso da exaltação e/ou crítica do materialismo, riqueza, luxúria, ostentação, futilidade, opulência, gratificação instantânea, superficialidade, aparência, do banal, do efêmero distanciando-se do sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente.
- Dimensão Musical 2 – Musicalidade baseada na intensidade, instrumentação amplificada, extroversão, vocalização vigorosa, pujante, com alto volume, força e movimentação rápida, em ritmo acelerado, voltada primeiramente para engajamento com a plateia em performance ao vivo, com 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, descrição de processos mentais internos, descrevendo o mundo como sobrenatural, místico, metafísico, figurado, lírico, poético, transcendente.
- Dimensão Musical 3 – Musicalidade baseada em acordes mais alegres e tons mais altos, com sonoridade menos emotiva e menos temperamental retratando 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, da busca incessante pela felicidade.

As dimensões foram rotuladas com descritores longos e complexos, visto que as funções apresentadas pelas classes semânticas e acústicas realizam discursos e musicalidade diferentes, com muitas facetas complementares. Sendo assim, a fim de condensar esses rótulos, propomos os seguintes rótulos sucintos, que capturam a essência discursiva de cada dimensão:

- Dimensão Musical 1 - Musicalidade para **dançar** retratando o **discurso materialista concreto e distanciando-se do místico**.
- Dimensão Musical 2 - Musicalidade **intensa** retratando o **discurso do drama pessoal e do sobrenatural**.
- Dimensão Musical 3 - Musicalidade **alegre** retratando o discurso **trágico-romântico**.

5. CONSIDERAÇÕES FINAIS

*Songs and sounds act as anchors. They fix for a moment, space and time*³⁰

Serge Lacasse

A pesquisa apresentada na presente tese teve como objetivo geral o desenvolvimento de uma descrição multimodal da música popular contemporânea em inglês, isto é, que envolvesse mais de um modo de expressão, no caso tanto o componente verbal (a letra da música) quanto o componente acústico (a sonoridade dos instrumentos). Além disso, o estudo pretendeu ser multidimensional, na medida em que cada canção foi mensurada em várias dimensões semânticas e acústicas, capazes de revelar o discurso presente no registro música em língua inglesa. Empregou também *big data*, pois foi baseada em uma grande amostra representativa da população de canções populares em inglês.

Para atingir tais objetivos, a presente pesquisa respondeu às seguintes perguntas: (1) Quais as dimensões semânticas do registro musical em língua inglesa? (2) Quais as dimensões sonoras do registro musical de gravações em língua inglesa? e (3) Quais as correlações entre as dimensões semânticas e acústicas no registro musical em língua inglesa?

Os resultados da Análise Multidimensional Semântica do corpus de estudo, composto por mais de 201 mil músicas em língua inglesa entre 1913 e 2018, responderam à questão 1 da presente pesquisa (Quais as dimensões semânticas do registro musical em língua inglesa?).

Werner (2021) e Delfino (2016) trabalharam com *corpus* de música e análise multidimensional, mas o que esses autores não levaram em consideração foi a análise lexical desse tipo de registro. Bértoli-Dutra (2014) foi a que mais se aproximou, ao fazer uma análise semântica (porém com um *corpus* e uma quantidade de variáveis menor do que utilizamos na presente tese). A autora encontrou um destaque para ações pessoais, emoções e sociedade e referência à música, o que também vemos na presente tese, onde os resultados da Análise Multidimensional Semântica das letras de música do corpus

³⁰ “Músicas e sons agem como âncoras. Eles te fixam num momento, espaço e tempo” (tradução nossa).

CoLiE apresentam quatro discursos bem distintos, marcadamente representando o materialismo (Dimensão Semântica 1), o existencialismo (Dimensão Semântica 2), o sobrenatural (Dimensão Semântica 3) e o trágico-romântico (Dimensão Semântica 4), com algumas complementaridades nessas dimensões. As Dimensões Semânticas 1 e 3 empregam muitas vezes textos descritivos para retratar os discursos do materialismo e do sobrenatural (respectivamente), enquanto que as Dimensões Semânticas 2 e 4, além de se valerem de elementos descritivos, estão associados com elementos narrativos – ou seja, ao contar um drama pessoal ou a tragédia romântica (discursos muito parecidos), as letras de música se valem muitas vezes tanto da descrição quanto da narração, simultaneamente.

Há uma correlação moderada entre o número de palavras (tokens) das letras de música com Dimensão Semântica 1 (referente ao discurso materialista), ou seja, há uma pequena tendência de que quanto maior o número de palavras em uma letra de música, maior o escore desse texto na Dimensão Semântica 1. A análise qualitativa mostrou que um gênero musical muito presente nessa dimensão é o hip-hop, cujas letras de música geralmente são mais longas do que outros gêneros musicais.

Excetuando-se a Dimensão Semântica 1, em que o hip-hop parece ser predominante entre as canções com maior escore, nas demais dimensões não parece haver um gênero predominante entre as canções com maior escore, segundo o que pôde ser observado na análise qualitativa, onde podemos notar que as músicas feitas para serem comercializadas dificultam as divisões em gêneros (com exceção da Dimensão Semântica 1). As Dimensões Semânticas 2, 3 e 4, todas as dimensões acústicas e todas as dimensões musicais possuem músicas de todos os gêneros.

Podemos afirmar que é muito difícil ouvirmos uma música que tenha atingido top hits de sucesso como o da Billboard³¹ ou ganhou algum Grammy³² e afirmarmos que essa música pertence ao gênero X ou Y. Há uma fluidez muito grande entre um gênero e outro, tanto na parte acústica quanto na escrita. A análise semântica nos mostrou isso, pois, ao fazermos a análise de mais de 200 mil músicas de todos os gêneros, elas se agruparam semanticamente em apenas quatro grupos, sem uma divisão distinta entre o

³¹ <https://www.billboard.com/charts/hot-100/>

³² <https://www.grammy.com/>

que é pop, rock, blues etc. O mesmo ocorreu com a análise acústica, onde mais de 91 mil músicas de todos os gêneros foram agrupadas em apenas três grupos.

Os resultados da análise multidimensional acústica do *corpus* de estudo, composto por mais de 95 mil músicas em língua inglesa entre 1913 e 2018, responderam à questão 2 da presente pesquisa (Quais as dimensões sonoras do registro musical em língua inglesa?). O que pudemos observar é que as músicas em língua inglesa apresentam três aspectos padrões bem distintos, os quais correspondem à: musicalidade intensa versus intimista (Dimensão Acústica 1), musicalidade para dançar versus para sentir (Dimensão Acústica 2) e musicalidade alegre versus triste (Dimensão Acústica 3).

Há uma alta correlação entre as Dimensões Acústicas 1 e 2, ou seja, a musicalidade intensa está altamente relacionada com a musicalidade para dançar, enquanto as músicas mais intimistas estão relacionadas com músicas feitas para sentir. Por outro lado, há uma baixa correlação negativa entre a Dimensão Acústica 3 e as outras duas anteriores, ou seja, musicalidade intensa e para dançar não quer dizer necessariamente uma música alegre, ao mesmo tempo que uma música intimista, feita para sentir, não é necessariamente uma música triste.

Os resultados da análise canônica do *corpus* de estudo responderam à questão 3 da presente pesquisa (Quais as correlações entre as dimensões semânticas e acústicas no registro musical em língua inglesa?). O que pudemos observar é que as músicas em língua inglesa apresentam três aspectos musicais bem distintos, representando a musicalidade para dançar retratando o discurso materialista concreto e distanciando-se do místico (Dimensão Musical 1), a musicalidade intensa retratando o discurso do existencialismo e do sobrenatural (Dimensão Musical 2) e a musicalidade alegre retratando o discurso trágico-romântico (Dimensão Musical 3).

Apesar desses resultados inovadores, em se tratando de pesquisa de base empírica, que usa big data e alia aspectos linguísticos e sonoros no registro música em língua inglesa, há pelo menos três direções em que a pesquisadora pretende continuar a desenvolver pesquisas futuras: a primeira, o desenvolvimento de uma tipologia textual da música popular contemporânea em inglês; a segunda, o estudo da variação diacrônica a fim de verificar até que ponto as dimensões identificadas relacionam-se com épocas específicas; a terceira, o desenvolvimento de software para ajudar a disseminar o conhecimento construído na pesquisa – por exemplo, um aplicativo de dispositivo móvel

em que o usuário selecionasse uma canção e o aplicativo retornasse suas características dimensionais.

Gostaria de terminar a presente tese relatando como tudo começou. A pesquisa se iniciou a partir de uma inquietação da pesquisadora sobre a categorização das músicas em gêneros sem levar em consideração a letra de música e a falta de estudos que considerassem a parte lexical, discursiva das músicas. Com o passar do tempo, fui percebendo que esse registro tão único, onde o acústico está tão intrinsecamente relacionado à letra, merecia um estudo amplo que levasse em conta a letra da música (aqui levado a cabo por meio da análise multidimensional semântica) e a sonoridade musical (aqui levado a cabo pela análise multidimensional acústica) e, conseqüentemente da união dessas duas modalidades, a verbal e a acústica (aqui levado a cabo por meio da análise de correlação canônica, da qual foram derivadas dimensões musicais). Podemos pensar que ainda existem muitos outros aspectos envolvidos que necessitam e merecem ser estudados, mas para a presente pesquisadora, letra e som se fazem muito presentes, pois, como afirma uma citação de um autor desconhecido, “alguns dias eu preciso da música e em outros eu preciso das letras”³³.

³³ “Some days I need the music and some days I need the lyrics”.

6. REFERÊNCIAS

ALMELA, A.; CANTOS GÓMEZ, P.; BERBER SARDINHA, T. Métodos multidimensionales basados en corpus del español. In: PARODI, G.; CANTOS GÓMEZ, P.; HOWE, L. **The Routledge Handbook of Spanish Corpus Linguistics**. Abingdon: Routledge, no prelo.

ALTSCHULER, G.C. **All shook up: How rock n roll changed America**. Oxford University Press, 2003.

ANDRESEN, L. **Battle notesmusic of the Vietnam War**. 2a edição, Savage Press, 2003.

ANDERON, R. "Reggae Music: A History and Selective Discography." **Notes**, vol. 61 no. 1, p. 206-214. *Project MUSE*, 2004, doi:10.1353/not.2004.0085.

ARCHER, D.; WILSON, A.; RAYSON, P. Introduction to the USAS category system. **Benedict project report**, 2002.

ARCHER, D.; RAYSON, P.; PIAO, S.; McENERY, T. Comparing the UCREL Semantic Annotation Scheme with lexicographical taxonomies. In G. Williams & S. Vessier (Eds.), **Proceedings of the 11th EURALEX (European Association for Lexicography) International Congress** (Euralex 2004) (Vol. III, pp. 817–827). Lorient: Université de Bretagne Sud, 2004.

AUSTIN, W. Hollywood Barn Dance: A Brief Survey of Country Music in Films. **Southern Quarterly; Hattiesburg** Vol. 22, Ed. 3: 111, Spring, 1984

BARBOSA, F. Um corpus do samba carioca para estudos lexicográficos e discursivos. In: SHEPHERD, T.; BERBER SARDINHA, T. e VEIRANO PINTO, M. **Caminhos da Linguística de Corpus**. Campinas, SP: Mercado de Letras, 2012. p. 247-270.

BENNETT, A. Towards a cultural sociology of popular music. **Journal of Sociology**, 44, n. 4, p. 419-432, 2008.

BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole, 2004. 410 p. p.

_____, T. **A corpus-based history of Applied Linguistics**. Apresentação de Trabalho. Associação Internacional de Linguística Aplicada (AILA), Rio de Janeiro, 2017(a).

_____, T. **Dimensions in collocation in American English**. Apresentação de Trabalho. Corpus Linguistics, Birmingham, 2017(b).

_____, T. Dimensions of variation across Internet registers. **International Journal of Corpus Linguistics**, 23, n. 2, p. 125-157, 2018.

_____, T. Lexicogrammar. In: CHAPELLE, C. **The Concise Encyclopedia of Applied Linguistics**. Hoboken, NJ: Wiley, p. 701-705, 2020.

_____, T. Discourse of academia from a multi-dimensional perspective. In: FRIGINAL, E. e HARDY, J. **The Routledge Handbook of Corpus Approaches to Discourse Analysis**. Abingdon: Routledge, p. 298-318, 2021.

_____, T. Corpus linguistics and the study of social media: a case study using multi-dimensional analysis. In: O'KEEFFE, A. e MCCARTHY, M. **The Routledge Handbook of Corpus Linguistics**. 2nd ed. New York, NY: Routledge, 2022.

_____, T. Going multimodal in corpus linguistics: The case of social media. Webinar apresentado na Queensland University:
<https://www.youtube.com/watch?v=eLlfoqaeREs>. Acesso em 20 de janeiro de 2022.

BERBER SARDINHA, T.; MADUREIRA, S. ; BRAIT, B.; SOUZA-E-SILVA, M. C. P. et al. Portal multimodal/multilíngue para o avanço da ciência aberta nas humanidades. **Cadernos de Linguística**, 2021.

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

BERBER SARDINHA, T.; KAUFFMANN, C.H.; ACUNZO, C. A multidimensional analysis of register variation in Brazilian Portuguese. **Corpora**, v.9, n.2, p. 239-271, 2014.

BERBER SARDINHA, T.; VEIRANO PINTO, M. American television and off screen registers: A corpus-based comparison. **Corpora**, 12, n. 1, p. 85-114, 2017.

BERBER SARDINHA, T.; VEIRANO PINTO, M. Dimensions of variation across American television registers. **International Journal of Corpus Linguistics**, 24, n. 1, p. 3-32, 2019a.

BERBER SARDINHA, T.; VEIRANO PINTO, M. (ed.). **Multi-Dimensional Analysis: Research Methods and Current Issues**. London: Bloomsbury Academic, 2019b.

BERBER SARDINHA, T. A historical characterization of American and Brazilian cultures based on lexical representations. In: *Corpora*, v.14, n.2, pp. 183-212, 2020(b)

BÉRTOLI DUTRA, P. **Linguagem da música popular anglo-americana de 1940-2009**. (Tese de Doutorado) - LAEL, PUCSP, São Paulo, SP. 2010.

<https://tedeantiga.pucsp.br/handle/handle/14145>

BERTOLI-DUTRA, P. Multi-dimensional analysis of pop songs. In: BERBER SARDINHA, T. e VEIRANO PINTO, M. **Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber**. Amsterdam/Philadelphia, PA: John Benjamins, p.149-175, 2014.

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

_____, D. Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. **Computational Linguistics** 19, p. 549-556, 1993.

_____, D. **Dimensions of Register Variation: A Cross-Linguistic Comparison**. Cambridge: Cambridge University Press, 1995. Disponível em: <http://dx.doi.org/10.1017/CBO9780511519871>.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus Linguistics - Investigating Language Structure and Use**. Cambridge: Cambridge University Press, 1998.

BIBER, D.; DAVIES, M.; JONES, J.K.; TRACY-VENTURA, N. Spoken and written register variation in Spanish: A multi-dimensional analysis. **Corpora**, v.1, n.1, p. 1-37, 2006.

BIBER, D.; KURJIAN, J. Towards a taxonomy of web registers and text types: a multi-dimensional analysis'. In: HUNDT, M. NESSELHAUF, N. and BIEWER, C. (eds) **Corpus Linguistics and the Web**, 109–32. Amsterdam and New York: Rodopi., 2007.

- BINDI, R.; CALZOLARI, N.; MONACHINI, M.; PIRRELI, V. Lexical knowledge acquisition from textual corpora: A multivariate statistic approach as an integration to traditional methodologies. In: **Proceedings, Seventh Annual Conference of the UW Centre for the New OED and Text Research**. Oxford, U.K., 1991
- BRADLEY, L. **Bass culture: When reggae was king**. Penguin Books. 2000
- BROWN, C.T. **The art of rock and roll**. 3a. edição. Englewood Cliffs, NJ, Prentice-Hall. 1992.
- CAMPBELL, M. **Popular music in America: The beat goes on**. Cengage Learning, 5a edição, 2018.
- COLLINS **English Dictionary**. 5th ed. Glasgow: Harper-Collins Publishers, 2001.
- DELFINO, M. C. N. Multidimensional analysis: number in Linguistics. **Cadernos de Linguística**, v. 2, n. 4, p. e 474, 11 Sep. 2021.
- DELFINO, M. C. N. **Uso de Música para o Ensino de Inglês como Língua Estrangeira em um Ambiente Baseado em Corpus**. (Dissertação de Mestrado) - LAEL, PUCSP., 2016. Disponível em: <https://tedeantiga.pucsp.br/handle/handle/18896>
- DELFINO, M.C.N.; FONSECA DE ARAÚJO, R.; BERBER SARDINHA, T. R. Revista Brasileira de Linguística Aplicada: multidimensões temáticas. In: FINATTO, M. J.B.; REBECHI, R. R.; SARMENTO, S.; BOCORNY, A. E. P. (Orgs.). **Linguística de Corpus: Perspectivas**. Porto Alegre: Instituto de Letras da UFRGS, 575p., 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,
- EARLY, G. **One nation under a groove**. 2a. edição. University of Michigan Press, 2004.
- ECHOLS, A. **Hot stuff: Disco and remaking of American Culture**. W.W. Norton & Company, 2010.
- EGBERT, J.; STAPLES, S. Doing Multi-Dimensional Analysis in SPSS, SAS, and R. In: BERBER SARDINHA, T. e VEIRANO PINTO, M. **Multi-dimensional analysis: Research methods and current issues**. London / New York: Bloomsbury / Continuum, 2019. p. 125-144.

EGBERT, J.; BIBER, D. Do all roads lead to Rome?: Modeling register variation with factor analysis and discriminant analysis. **Corpus Linguistics and Linguistic Theory**, v. 14, n.2, p. 233-273, set. de 2018.

EYERMAN, R.; JAMISON, A. Social movements and cultural transformation: popular music in the 1960s. **Media, Culture & Society**. 17(3):449-468, 1995. Disponível em: doi:10.1177/016344395017003006

FRIGINAL, E.; HARDY, J. A. Conducting Multi-Dimensional analysis using SPSS. In: BERBER SARDINHA, T. e VEIRANO PINTO, M. **Multi- 19 Dimensional Analysis, 25 years on: A Tribute to Douglas Biber**. Amsterdam/Philadelphia, PA: John Benjamins, 2014. p. 298-316.

FRITH, S. “Rock and the Politics of Memory.” **Social Text**, no. 9/10, Duke University Press, 1984, pp. 59–69. Disponível em: <https://doi.org/10.2307/466535>.

GABBARD, K. **Jazz among the discourses**. 1a edição. Duke University Press. Project MUSE muse.jhu.edu/book/70491, 1995.

GEORGE, N. **Thriller: The musical life of Michael Jackson**. Da Capo Press, 2010.

GLEZ-PEÑA, D., LOURENÇO, A., LÓPEZ-FERNÁNDEZ, H. REBOIRO-JATO, M. FDEZ-RIVEROLA, F. Web scraping technologies in an API world, **Briefings in Bioinformatics**, Volume 15, Issue 5, Setembro de 2014, pp. 788–797. Disponível em: <https://doi.org/10.1093/bib/bbt026>

GORDY, B. **To be loved: The music, the magic, the memories of motown**. Rosetta Books, 1994.

GRANDE, S.V.L. **O impacto do rock no comportamento do jovem**. Tese de Doutorado, UNESP Araraquara, Programa de Pós Graduação em Sociologia, 2006. <https://repositorio.unesp.br/handle/11449/106266>

GROSSBERG, L. **We gotta get out of this place: popular conservatism and postmodern culture**. New York, Routledge, 1992.

HONGPANARAK, T.; MONGKOLNAVIN, J. **A Study of Relationship Between Music Streaming Behavior and Big Five Personality Traits of Spotify Users**. Comunicação apresentada em: IAIT2021: The 12th Conference on Advances in

Information Technology. Disponível em: <https://doi.org/10.1145/3468784.3469854>, 2021

KAUFFMANN, C.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. <https://tede2.pucsp.br/handle/handle/23128>

LACASSE, S. Slave to the Supradiegetic Rhythm: A Micro-Rhythmic Analysis of Creaky Voice in Sia's 'Breath Me. In: DANIELSEN, A. **Musical Rhythm in the Age of Digital Reproduction**, 141–55. Aldershot: Ashgate (2010).

LEECH, G. Corpora and theories of linguistic performance. In: J. SVARTVIK (org.). **Directions in Corpus Linguistics**. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991 Berlin, New York: De Gruyter, 1992.

LULL, J. (org.) **Popular Music and Communication**. 2a edição. Newbury Park, CA, Sage, 1992.

MARCUS, G. **In the Fascist Bathroom: Punk in pop music 1977-1992**. Harvard University Press, 1993.

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. <https://tede2.pucsp.br/handle/handle/21448>

MAYER, R.; NEUMAYER, R.; RAUBER, A. Rhyme and style features for musical genre classification by song lyrics. In: BELLO, J.P.; CHEW, E.; TURNBULL, D. (ed.) **ISMIR 2008: Proceedings of the 9th International Conference of Music**, p. 337-342

MCENERY, T.; WILSON, A. **Corpus linguistics**. Edinburgh: Edinburgh University Press, 1996. 209 pp.

MIDDLETON, R. Popular music analysis and Musicology: Bridging the gap. **Popular Music**, 12, n. 2, p. 177-190, 1993.

- MILANEZ, A.A.Z.; KAUFFMANN, C.H.; DELFINO, M.C.N.; BERTOLI, P.P.; DÁRTORA, T. **As dimensões do twitter contra e a favor das manifestações do 7 de setembro de 2021**. Comunicação apresentada no InPLA, 2021.
- MOORE, A. F. **Song means: Analyzing and interpreting recorded popular song**. Farnham: Ashgate, 2021.
- MULLEN, J. **Popular song in the First World War: An international perspective**. Routledge, 2020, 270 pp.
- MURPHEY, T. The song stuck in my head phenomenon: A melodic din in the lad?. **System**, Volume 18, Issue 1, p. 53-64, 1990, ISSN 0346-251X, disponível em: [https://doi.org/10.1016/0346-251X\(90\)90028-4](https://doi.org/10.1016/0346-251X(90)90028-4)
- NILE, N.N. **Rock On Almanac**, 2nd ed., New York: HarperCollins Publishers, 1992.
- OCHI, V.; ESTRADA, R. GAJI, T.; GADEA, W.; DUONG, E. **Spotify Danceability and Popularity Analysis using SAP**. In: <https://arxiv.org/abs/2108.02370>. Acesso em 12 de janeiro de 2022.
- OLWIG, K.R. The ‘British invasion’: the ‘new’ cultural geography and beyond cultural geographies. **SAGE Journals**, 17(2):175-179, 2010. Disponível em: doi:10.1177/1474474010363854
- PALHA, A.P. **Fãs brasileiros de K-pop: Um estudo sobre aculturação de consumo**. Dissertação de Mestrado. Universidade Federal de Pernambuco. Programa de Pós-Graduação em Administração, 2021.
<https://repositorio.ufpe.br/handle/123456789/40186>
- PANDA, R.; REDINHO, H.; GONÇALVES, C.; MALHEIRO, R.; PAIVA, R.P. **How Does the Spotify API Compare to the Music Emotion Recognition State-of-the-Art?** In: Proceedings of the 18th Sound and Music Computing Conference (SMC 2021). <http://hdl.handle.net/10316/95161>. Disponível em: <https://doi.org/10.5281/zenodo.5045100>, 2021
- PAVLIK, K. **Understanding + classifying genres using Spotify audio features**. 2019: <https://www.kaylinpavlik.com/classifying-songs-genres/> Acesso em 06 de fevereiro de 2022

PINARBAŞI, F. Demystifying musical preferences at Turkish music market through audio features of Spotify charts. **Turkish Journal of Marketing**, 4(3), 264-279.

<https://doi.org/10.30685/tujom.v4i3.62>, 2019.

RAYSON, P. **Wmatrix: A web-based corpus processing environment**. Computing Department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/>, 2007. Acesso em 28 de janeiro de 2022.

RIPANI, R.J. **The new blue music: Changes in Rhythm & Blues, 1950-1999**.

University Press of Mississippi, 2006.

ROCKBIZZ. **Há 40 Anos a MTV Anunciava o Fim do Rádio**. In:

<https://www.rockbizz.com.br/ha-40-anos-a-mtv-anunciava-o-fim-do-radio/> Acesso em 20 de novembro de 2021.

ROMEIRO Y.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) - Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020.

<https://sapientia.pucsp.br/handle/handle/23462>

SALSBURY T., CROSSLEY, S.A., MCNAMARA, DS. Psycholinguistic word information in second language oral discourse. **Second Language Research**: 27(3):343-360, 2011. Disponível em: doi:10.1177/0267658310395851

SEABROOK, J. **The Song Machine: Inside the Hit Factory**. W. W. Norton & Company, New York – London, 2015.

SINCLAIR, J. MCH. **Corpus, Concordance, Collocation**. Oxford, New York: Oxford University Press, xviii, 179 p., 1991.

SINCLAIR, J. MCH.; JONES, S.; DALEY, R. English Lexical Studies: **The OSTI Report**. London/New York: Continuum, 1970/2004. 2-138 p.

SHUKER, R. **Vocabulário de música pop**. Routledge Ltd., 1998.

STRATTON, J. Why doesn't anybody write anything about glam rock? **Aust. J. Cultural Studies**, 4:1, 1986.

TABACHNICK, B.G.; FIDELL, L.S. **Using multivariate statistics**. Boston: Pearson Education, 2007.

THACKRAY, J. In Search of Nirvana: Why Nirvana: The True Story Could Never Be “True”. **Popular Music and Society**, 38:2, pages 194-207, 2015.

VEIGA, A.T. **As dimensões da fé: sete religiões mundiais em uma análise multidimensional lexical**. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) - Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem da Pontifícia Universidade Católica de São Paulo, São Paulo, 2021. <https://repositorio.pucsp.br/jspui/handle/handle/23988>

VEIRANO PINTO, M. Dimensions of variation in North American movies. In: BERBER SARDINHA, T. e VEIRANO PINTO, M. **Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber**. Amsterdam/Philadelphia, PA: John Benjamins, 2014. p. 109-149.

VLADI, N. (2010). **O negócio da música – como os gêneros musicais articulam estratégias de comunicação para o consumo cultural**. Comunicação apresentada para o GP Comunicação e Culturas Urbanas (DT6), X Encontro dos Grupos de Pesquisas em Comunicação, evento componente do XXXIII Congresso Brasileiro de Ciências da Comunicação.

WALL, T. **Studying Popular Music Culture**. Newbury Park: Sage, 2013.

WELSH, R. Rock’n’roll and social change. **History Today**, pp. 32-39, Feb. 1990

WERNER, V. Catchy and conversational? A register analysis of pop lyrics. **Corpora**, 16, n. 2, p. 237-270, 2021

_____, V. Love is all around: A corpus-based study of pop music lyrics. **Corpora**, 7, n. 1, p. 19-50, 2012.

WILKINSON, D.; WORLEY, M.; STREET, J. ‘I Wanna See Some History’: Recent Writing on British Punk. **Contemporary European History**, 26:2, pages 397-411, 2017.

WILLIAMSON, V. **You are the music: how music reveals what it means to be human**. Icon Books, Ltd, 2014.

ZUPPARDI, M.C. **Collocation dimensions in academic English**. 243 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.

<https://sapientia.pucsp.br/handle/handle/23642>

7. ANEXOS

ANEXO 1 - PROGRAMA SAS PARA ANÁLISE DE *CLUSTERS* DAS VARIÁVEIS

```
/* BEGINNING PART 1 */
/* ==== EDIT BELOW =====*/

%let project = claudia_semantic ;

%let myfolder = &project ;

%let sasusername = u50418389 ;

%let whereisit = /home/&sasusername ; /* online */

libname user "&whereisit/&myfolder";

options fmtsearch=(user work library);

/* END PART 1 */
/* BEGINNING PART 2 */

OPTIONS VALIDVARNAME=ANY;
FILENAME IN "&whereisit/&myfolder/&project..tsv";
PROC IMPORT OUT= &project
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

/* data &project (drop=A1 A3 A6 A7 A8 Z0-Z9 E2 I4 K3 N3 N4 O1 T1 W5 X1 X3 VAR111
); */

data &project (drop=A1 A3 A6 A7 A8 Z0-Z9 E2 I4 K3 N3 N4 O1 T1 W5 VAR111 );
set &project;
run;

/* remove lines that are all zeros */

data temp (DROP= file folder );
set &project ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;
```

```

proc sql noprint ;
  select _name_ into :names separated by ' + ' from rot ;
quit;

data &project (drop= total);
set &project;
total= &names ;
if total > 0 ;
run;

/* calculate TTR */
OPTIONS VALIDVARNAME=ANY;
FILENAME IN "&whereisit/&myfolder/types_tokens.tsv";
PROC IMPORT OUT= typetoken
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

data typetoken ;
set typetoken;
if tokens > 0;
tr = types/tokens ;
run;

OPTIONS VALIDVARNAME=ANY;
FILENAME IN "&whereisit/&myfolder/language.tsv";
PROC IMPORT OUT= language
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

proc sort data=&project ; by file; run;
proc sort data=typetoken ; by file; run;
proc sort data=language ; by file; run;

data temp;
merge &project (in=a) typetoken (in=b) ;
by file;
if (a and b) then output;
run;

data temp2;
merge temp (in=a) language (in=b) ;
by file;
if (a and b) then output;

```

```

run;

/* keep songs with at least x% english words */
data &project ;
set temp2 ;
if english > 50 ;
run;

/* norm the data */

data temp (DROP= file folder types english );
set &project ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;
proc sql noprint ;
  select _name_ into :names separated by ' ' from rot ;
quit;

data normed;
set &project ;
array v &names ;
do over v ;
  v = ( v / tokens ) * 1000;
end ;
run;

/* variable clustering */
/* https://www.lexjansen.com/nesug/nesug11/sa/sa08.pdf */

data temp (DROP= file folder tokens types english );
set normed ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint ;
  select _name_ into :names separated by ' ' from rot ;
quit;

PROC VARCLUS
DATA = normed
MAXEIGEN = .7
MAXCLUSTERS = 30
SHORT
HI;
VAR &names ;
ODS OUTPUT
RSQUARE = OUTPUT_DATA;
RUN;

```

```

/* select variables from varclus output */
data varclus_selected (keep= variable);
set output_data;
if NumberOfClusters = 30;
if OwnCluster >= .3 ;
run;

```

ANEXO 2 - PROGRAMA SAS PARA ANÁLISE FATORIAL SEMÂNTICA

```

/* NORMING, VARCLUS FEATURE SELECTION, RANKING, POLYCHOR, NO Z-
SCORE */
/* TOP EIGEN = ?? LONGER SONGS = ?? */

/* BEGINNING PART 1 */
/* ==== EDIT BELOW =====*/

%let project = claudia_semantic ;

%let myfolder = &project ;

%let sasusername = u50418389 ;

%let whereisit = /home/&sasusername ; /* online */

libname user "&whereisit/&myfolder";

options fmtsearch=(user work library);

/* enter number of factors to extract */
%let extractfactors = 4 ;

/* enter factor variable names */
%let factorvars = fac1-fac4 ;

/* enter min loading cutoff */
%let minloading = .3 ;

/* enter min communality cutoff */
%let communalcutoff = .15 ;

/* enter the top n most freq vars cutoff ; to retain all vars, enter high value */
%let topn = 100 ;

/* END PART 1 */
/* BEGINNING PART 2 */

OPTIONS VALIDVARNAME=ANY;

```

```

FILENAME IN "&whereisit/&myfolder/&project..tsv";
PROC IMPORT OUT= &project
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

/* data &project (drop=A1 A3 A6 A7 A8 Z0-Z9 E2 I4 K3 N3 N4 O1 T1 W5 X1 X3 VAR111
); */

data &project (drop=A1 A3 A6 A7 A8 Z0-Z9 E2 I4 K3 N3 N4 O1 T1 W5 VAR111 );
set &project;
run;

/* remove lines that are all zeros */

data temp (DROP= file folder );
set &project ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint ;
  select _name_ into :names separated by ' + ' from rot ;
quit;

data &project (drop= total);
set &project;
total= &names ;
if total > 0 ;
run;

/* calculate TTR */

OPTIONS VALIDVARNAME=ANY;
FILENAME IN "&whereisit/&myfolder/types_tokens.tsv";
PROC IMPORT OUT= typetoken
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

data typetoken ;
set typetoken;
if tokens > 0;
  ttr = types/tokens ;
run;

```

```

OPTIONS VALIDVARNAME=ANY;
FILENAME IN "&whereisit/&myfolder/language.tsv";
PROC IMPORT OUT= language
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

proc sort data=&project ; by file; run;
proc sort data=typetoken ; by file; run;
proc sort data=language ; by file; run;

data temp;
merge &project (in=a) typetoken (in=b) ;
by file;
if (a and b) then output;
run;

data temp2;
merge temp (in=a) language (in=b) ;
by file;
if (a and b) then output;
run;

/* keep songs with at least x% english words */
data &project ;
set temp2 ;
if english > 50 ;
run;

/* norm the data */
data temp (DROP= file folder ttr tokens types english );
set &project ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint ;
  select _name_ into :names separated by ' ' from rot ;
quit;

data normed;
set &project ;
array v &names ;
do over v ;
  v = ( v / tokens ) * 1000;
end ;
run;

```

```

/* use variables selected by VARCLUS in a separate program */

proc sql ;
  select variable into :names separated by ' ' from varclus_selected ;
quit;

data normed_topn (keep= &names folder file english ttr types tokens );
set normed ;
run;

proc datasets library=user nolist;
delete
&project rot temp temp2 temp3 want long long_marked
out out2 corr pearson polychor ranked;
run;

/* BEGINNING PART 3 */

/* rank the data */

/* turn zeros to missing */

/* speed up by picking a single line of data to rotate */
data temp (DROP= file folder english types );
set normed_topn ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint;
  select _name_ into :names separated by ' ' from rot ;
quit;

data normed_topn ;
set normed_topn ;
array v &names ;
do i=1 to dim(v);
  if v[i]=0 then v[i]=.;
end;
drop i;
run;

/*https://www.listendata.com/2016/01/proc-rank.html */
OPTIONS VALIDVARNAME=ANY;
proc rank data = normed_topn groups = 5 out = ranked ties = high ;
var &names ;
run;

/* ranks start with 0; add 1 */
data ranked ;
set ranked;
array v &names ;
do over v ;

```

```

v = v + 1;
end ;
run;

/* turn missing values to zeros */
/* https://communities.sas.com/t5/SAS-Procedures/How-to-set-all-missing-values-to-zero-
for-all-variables/td-p/41478 */

proc stdize data = ranked out=ranked reponly missing=0; run;

PROC EXPORT
DATA= user.ranked
DBMS=TAB
OUTFILE="&whereisit/&myfolder/ranked.tsv"
REPLACE;
RUN;

/* compute correlations */

data temp (DROP= file folder english types tokens);
set ranked ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint;
select _name_ into :names separated by ' ' from rot ;
quit;

/* SLOW */
proc corr data = ranked outplc = polychor polychoric noprint;
var &names ;
run;

/* turn missing correlation values to zeros */
proc stdize data = polychor out=polychor reponly missing=0; run;

/* export to check if correlation matrix is singular: */

PROC EXPORT
DATA= user.polychor
DBMS=TAB
OUTFILE="&whereisit/&myfolder/corr.tsv"
REPLACE;
RUN;

/* END PART 8 */
/* BEGINNING PART 9 */

/* number of observations IN THE DATA */
data _NULL_;
if 0 then set normed_topn nobs=n;

```

```

        call symputx('nobs',n);
        stop;
run;
%put nobs=&nobs ;

/* get variable list for factor */
/* speed up by picking a single line of data to rotate */
data temp (DROP= file folder types tokens english );
set ranked ;
if _n_ <=1 ;
run;

proc transpose data=temp out= rot ; run;

proc sql noprint;
  select _name_ into :names separated by ' ' from rot ;
quit;

/*BEGINNING PART */

/* unrotated, before dropping low communalities */

proc datasets library=user nolist;
delete
fout;
run;

ODS EXCLUDE NONE;
proc factor fuzz=0.3 data= polychor (type=corr) OUTSTAT= fout NOPRINT
method=principal
plots=scree
mineigen=1
reorder
heywood
nfactors=100
nobs=&nobs; /* specify number of obs because this is missing from a corr matrix */
var &names ;
run;

/* BEGINNING PART 10 */

**** find low communalities ****/

data fout2;
  set fout (where=( _TYPE_="COMMUNAL"));
run;

proc transpose data=fout2 out=communal; id _TYPE_; run;

/* list vars to drop */
proc sql ;

```

```

select _name_ into :lowcomm separated by ' ' from communal
  where communal < &communalcutoff ;
quit;

/* list vars to keep */

ODS EXCLUDE NONE ;
proc sql ;
  select _name_ into :highcomm separated by ' ' from communal
    where communal >= &communalcutoff ;
quit;

/* save communalities to spreadsheet */

PROC SORT data=communal (keep= _name_ communal); BY communal ; RUN;

PROC EXPORT
  DATA= USER.communal
  DBMS=TAB
  OUTFILE="&whereisit/&myfolder/communalities.tsv"
  REPLACE;
RUN;

/* END PART 11 */
/* BEGINNING PART 12 */

/* scree plot */

data fout2;
  set fout (where=( _TYPE_="EIGENVAL"));
run;

proc transpose data=fout2 out= fout3 (drop = _NAME_);
id _TYPE_;
run;

data fout4 ;
set fout3 ;
factor = _n_;
if factor <= 20 ;
run;

/* delete the scree files */
FILENAME temp "/home/&sasusername/&myfolder/scree_1.png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

FILENAME temp "/home/&sasusername/&myfolder/scree_2.png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

%macro create(howmany);

```

```

%do i=1 %to &howmany;

FILENAME temp "/home/&sasusername/&myfolder/scree_1&i..png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

FILENAME temp "/home/&sasusername/&myfolder/scree_2&i..png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

%end;
%mend create;
%create(9)
;

/* create the scree files */

ods listing gpath="&whereisit/&myfolder/";
ods graphics / imagename="scree_1" imagefmt=png;
title "Scree plot";
proc sgplot data= fout4 ;
series x=factor y=EIGENVAL / datalabel=EIGENVAL;
run;
title;

ods listing gpath="&whereisit/&myfolder/";
ods graphics / imagename="scree_2" imagefmt=png;
title "Scree plot";
proc sgplot data= fout4 ;
series x=factor y=EIGENVAL / datalabel=factor;
run;
title;

/* determine the number of factors and enter number in the header of this program */

/* END PART 12 */
/* BEGINNING PART 13 */

/* rotated w/o low communalities */
/* ignore error 'matrix is singular' because results will be output anyway */

proc datasets library=user nolist;
delete
rotatedfinal ;
run;

proc factor fuzz=0.3 data= polychor (type=corr) OUTSTAT= rotatedfinal NOPRINT
method=principal
msa
mineigen=0
priors=max /* max = matrix is singular */
nfactors= &extractfactors

```

```

rotate=promax
heywood
nobs=&nobs; /* specify number of obs because this is missing from a corr matrix */
var &highcomm ;
run;
/* BEGINNING PART 14 */

/* loadings table */

/*

https://stats.idre.ucla.edu/sas/output/factor-analysis/
Rotated Factor Pattern – This table contains the rotated factor loadings, which are the
correlations between the variable and the factor. Because these are correlations, possible
values range from -1 to +1.
in the outstat data file, the rotated factor pattern appears as PREROTAT. The standardized
regression coefficients appear as PATTERN.
Use PREROTAT in the outstat data file.

https://documentation.sas.com/?docsetId=statug&docsetTarget=statug\_factor\_details02.htm&docsetVersion=15.1&locale=en

PREROTAT: prerotated factor pattern.
PATTERN: factor pattern. (regression coefficients)

PREROTAT: prerotated factor pattern. => Stat.Factor.OrthRotFactPat
PATTERN: factor pattern. => Stat.Factor.ObliqueRotFactPat

*/

OPTIONS VALIDVARNAME=ANY;
data rotated2;
  set rotatedfinal (where=( _TYPE_="PREROTAT"));
run;

proc transpose data=rotated2 out= rotated2 ;
id _NAME_ ;
run;

OPTIONS VALIDVARNAME=ANY;
data rotated3;
  set rotated2;
  loaded = 0 ;
  if abs(factor1) > abs(factor2)
    AND abs(factor1) > abs(factor3)
    AND abs(factor1) > abs(factor4)
    AND abs(factor1) > abs(factor5)
    AND abs(factor1) > abs(factor6)
    AND abs(factor1) > abs(factor7)
    AND abs(factor1) > abs(factor8)
    AND abs(factor1) > abs(factor9)
    AND factor1 > 0 AND abs(factor1) >= "&minloading" then do; factor = 'fac1'; pole = 1;
loaded = 1; end ;

```

```

else if abs(factor2) > abs(factor1)
  AND abs(factor2) > abs(factor3)
  AND abs(factor2) > abs(factor4)
  AND abs(factor2) > abs(factor5)
  AND abs(factor2) > abs(factor6)
  AND abs(factor2) > abs(factor7)
  AND abs(factor2) > abs(factor8)
  AND abs(factor2) > abs(factor9)
  AND factor2 > 0 AND abs(factor2) >= "&minloading" then do; factor = 'fac2'; pole = 1;
loaded = 1; end ;
else if abs(factor3) > abs(factor1)
  AND abs(factor3) > abs(factor2)
  AND abs(factor3) > abs(factor4)
  AND abs(factor3) > abs(factor5)
  AND abs(factor3) > abs(factor6)
  AND abs(factor3) > abs(factor7)
  AND abs(factor3) > abs(factor8)
  AND abs(factor3) > abs(factor9)
  AND factor3 > 0 AND abs(factor3) >= "&minloading" then do; factor = 'fac3'; pole = 1;
loaded = 1; end ;

else if abs(factor4) > abs(factor1)
  AND abs(factor4) > abs(factor2)
  AND abs(factor4) > abs(factor3)
  AND abs(factor4) > abs(factor5)
  AND abs(factor4) > abs(factor6)
  AND abs(factor4) > abs(factor7)
  AND abs(factor4) > abs(factor8)
  AND abs(factor4) > abs(factor9)
  AND factor4 > 0 AND abs(factor4) >= "&minloading" then do; factor = 'fac4'; pole = 1;
loaded = 1; end ;

else if abs(factor5) > abs(factor1)
  AND abs(factor5) > abs(factor2)
  AND abs(factor5) > abs(factor3)
  AND abs(factor5) > abs(factor4)
  AND abs(factor5) > abs(factor6)
  AND abs(factor5) > abs(factor7)
  AND abs(factor5) > abs(factor8)
  AND abs(factor5) > abs(factor9)
  AND factor5 > 0 AND abs(factor5) >= "&minloading" then do; factor = 'fac5'; pole = 1;
loaded = 1; end ;

else if abs(factor6) > abs(factor1)
  AND abs(factor6) > abs(factor2)
  AND abs(factor6) > abs(factor3)
  AND abs(factor6) > abs(factor4)
  AND abs(factor6) > abs(factor5)
  AND abs(factor6) > abs(factor7)
  AND abs(factor6) > abs(factor8)
  AND abs(factor6) > abs(factor9)
  AND factor6 > 0 AND abs(factor6) >= "&minloading" then do; factor = 'fac6'; pole = 1;
loaded = 1; end ;

```

```

else if abs(factor7) > abs(factor1)
  AND abs(factor7) > abs(factor2)
  AND abs(factor7) > abs(factor3)
  AND abs(factor7) > abs(factor4)
  AND abs(factor7) > abs(factor5)
  AND abs(factor7) > abs(factor6)
  AND abs(factor7) > abs(factor8)
  AND abs(factor7) > abs(factor9)
  AND factor7 > 0 AND abs(factor7) >= "&minloading" then do; factor = 'fac7'; pole = 1;
loaded = 1; end ;

else if abs(factor8) > abs(factor1)
  AND abs(factor8) > abs(factor2)
  AND abs(factor8) > abs(factor3)
  AND abs(factor8) > abs(factor4)
  AND abs(factor8) > abs(factor5)
  AND abs(factor8) > abs(factor6)
  AND abs(factor8) > abs(factor7)
  AND abs(factor8) > abs(factor9)
  AND factor8 > 0 AND abs(factor8) >= "&minloading" then do; factor = 'fac8'; pole = 1;
loaded = 1; end ;

else if abs(factor9) > abs(factor1)
  AND abs(factor9) > abs(factor2)
  AND abs(factor9) > abs(factor3)
  AND abs(factor9) > abs(factor4)
  AND abs(factor9) > abs(factor5)
  AND abs(factor9) > abs(factor6)
  AND abs(factor9) > abs(factor7)
  AND abs(factor9) > abs(factor8)
  AND factor9 > 0 AND abs(factor9) >= "&minloading" then do; factor = 'fac9'; pole = 1;
loaded = 1; end ;

/* negative values */

else if abs(factor1) > abs(factor2)
  AND abs(factor1) > abs(factor3)
  AND abs(factor1) > abs(factor4)
  AND abs(factor1) > abs(factor5)
  AND abs(factor1) > abs(factor6)
  AND abs(factor1) > abs(factor7)
  AND abs(factor1) > abs(factor8)
  AND abs(factor1) > abs(factor9)
  AND factor1 < 0 AND abs(factor1) >= "&minloading" then do; factor = 'fac1'; pole = -1;
loaded = 1; end ;

else if abs(factor2) > abs(factor1)
  AND abs(factor2) > abs(factor3)
  AND abs(factor2) > abs(factor4)
  AND abs(factor2) > abs(factor5)
  AND abs(factor2) > abs(factor6)
  AND abs(factor2) > abs(factor7)
  AND abs(factor2) > abs(factor8)
  AND abs(factor2) > abs(factor9)

```

```
    AND factor2 < 0 AND abs(factor2) >= "&minloading" then do; factor = 'fac2'; pole = -1;
loaded = 1; end ;
```

```
else if abs(factor3) > abs(factor1)
    AND abs(factor3) > abs(factor2)
    AND abs(factor3) > abs(factor4)
    AND abs(factor3) > abs(factor5)
    AND abs(factor3) > abs(factor6)
    AND abs(factor3) > abs(factor7)
    AND abs(factor3) > abs(factor8)
    AND abs(factor3) > abs(factor9)
    AND factor3 < 0 AND abs(factor3) >= "&minloading" then do; factor = 'fac3'; pole = -1;
loaded = 1; end ;
```

```
else if abs(factor4) > abs(factor1)
    AND abs(factor4) > abs(factor2)
    AND abs(factor4) > abs(factor3)
    AND abs(factor4) > abs(factor5)
    AND abs(factor4) > abs(factor6)
    AND abs(factor4) > abs(factor7)
    AND abs(factor4) > abs(factor8)
    AND abs(factor4) > abs(factor9)
    AND factor4 < 0 AND abs(factor4) >= "&minloading" then do; factor = 'fac4'; pole = -1;
loaded = 1; end ;
```

```
else if abs(factor5) > abs(factor1)
    AND abs(factor5) > abs(factor2)
    AND abs(factor5) > abs(factor3)
    AND abs(factor5) > abs(factor4)
    AND abs(factor5) > abs(factor6)
    AND abs(factor5) > abs(factor7)
    AND abs(factor5) > abs(factor8)
    AND abs(factor5) > abs(factor9)
    AND factor5 < 0 AND abs(factor5) >= "&minloading" then do; factor = 'fac5'; pole = -1;
loaded = 1; end ;
```

```
else if abs(factor6) > abs(factor1)
    AND abs(factor6) > abs(factor2)
    AND abs(factor6) > abs(factor3)
    AND abs(factor6) > abs(factor4)
    AND abs(factor6) > abs(factor5)
    AND abs(factor6) > abs(factor7)
    AND abs(factor6) > abs(factor8)
    AND abs(factor6) > abs(factor9)
    AND factor6 < 0 AND abs(factor6) >= "&minloading" then do; factor = 'fac6'; pole = -1;
loaded = 1; end ;
```

```
else if abs(factor7) > abs(factor1)
    AND abs(factor7) > abs(factor2)
    AND abs(factor7) > abs(factor3)
    AND abs(factor7) > abs(factor4)
    AND abs(factor7) > abs(factor5)
    AND abs(factor7) > abs(factor6)
    AND abs(factor7) > abs(factor8)
```

```

AND abs(factor7) > abs(factor9)
AND factor7 < 0 AND abs(factor7) >= "&minloading" then do; factor = 'fac7'; pole = -1;
loaded = 1; end ;

else if abs(factor8) > abs(factor1)
AND abs(factor8) > abs(factor2)
AND abs(factor8) > abs(factor3)
AND abs(factor8) > abs(factor4)
AND abs(factor8) > abs(factor5)
AND abs(factor8) > abs(factor6)
AND abs(factor8) > abs(factor7)
AND abs(factor8) > abs(factor9)
AND factor8 < 0 AND abs(factor8) >= "&minloading" then do; factor = 'fac8'; pole = -1;
loaded = 1; end ;

else if abs(factor9) > abs(factor1)
AND abs(factor9) > abs(factor2)
AND abs(factor9) > abs(factor3)
AND abs(factor9) > abs(factor4)
AND abs(factor9) > abs(factor5)
AND abs(factor9) > abs(factor6)
AND abs(factor9) > abs(factor7)
AND abs(factor9) > abs(factor8)
AND factor9 < 0 AND abs(factor9) >= "&minloading" then do; factor = 'fac9'; pole = -1;
loaded = 1; end ;
run;

data rotated4 ; set rotated3 ; if loaded = 1; run; quit;

/* labeling: https://stats.idre.ucla.edu/sas/modules/labeling/ */
PROC FORMAT library=user ;
VALUE $semfeaturelabels
"A1" = "General And Abstract Terms (A1)"
"A2" = "Affect (A2)"
"A3" = "Being (A3)"
"A4" = "Classification (A4)"
"A5" = "Evaluation (A5)"
"A6" = "Comparing (A6)"
"A7" = "Probability (A7)"
"A8" = "Seem (A8)"
"A9" = "Getting and giving; possession (A9)"
"B1" = "Anatomy and physiology (B1)"
"B2" = "Health and disease (B2)"
"B3" = "Medicines and medical treatment (B3)"
"B4" = "Cleaning and personal care (B4)"
"B5" = "Clothes and personal belongings (B5)"
"C1" = "Arts and crafts (C1)"
"E1" = "Emotional Actions, States And Processes General (E1)"
"E2" = "Liking (E2)"
"E3" = "Calm/Violent/Angry (E3)"
"E4" = "Happiness and Contentment (E4)"
"E5" = "Bravery and Fear (E5)"
"E6" = "Worry and confidence (E6)"
"F1" = "Food (F1)"

```

"F2" = "Drinks and alcohol (F2)"
 "F3" = "Smoking and non-medical drugs (F3)"
 "F4" = "Farming & Horticulture (F4)"
 "G1" = "Government and Politics (G1)"
 "G2" = "Crime, law and order (G2)"
 "G3" = "Warfare, defence and the army; weapons (G3)"
 "H1" = "Architecture, houses and buildings (H1)"
 "H2" = "Parts of buildings (H2)"
 "H3" = "Areas around or near houses (H3)"
 "H4" = "Residence (H4)"
 "H5" = "Furniture and household fittings (H5)"
 "I1" = "Money generally (I1)"
 "I2" = "Business (I2)"
 "I3" = "Work and employment (I3)"
 "I4" = "Industry (I4)"
 "K1" = "Entertainment generally (K1)"
 "K2" = "Music and related activities (K2)"
 "K3" = "Recorded sound (K3)"
 "K4" = "Drama, the theatre and show business (K4)"
 "K5" = "Sports and games generally (K5)"
 "K6" = "Children's games and toys (K6)"
 "L1" = "Life and living things (L1)"
 "L2" = "Living creatures: animals, birds, etc. (L2)"
 "L3" = "Plants (L3)"
 "M1" = "Moving, coming and going (M1)"
 "M2" = "Putting, pulling, pushing, transporting (M2)"
 "M3" = "Vehicles and transport on land (M3)"
 "M4" = "Sailing, swimming, etc. (M4)"
 "M5" = "Flying and aircraft (M5)"
 "M6" = "Location and direction (M6)"
 "M7" = "Places (M7)"
 "M8" = "Stationary (M8)"
 "N1" = "Numbers (N1)"
 "N2" = "Mathematics (N2)"
 "N3" = "Measurement (N3)"
 "N4" = "Linear order (N4)"
 "N5" = "Quantities (N5)"
 "N6" = "Frequency (N6)"
 "O1" = "Substances and materials generally (O1)"
 "O2" = "Objects generally (O2)"
 "O3" = "Electricity and electrical equipment (O3)"
 "O4" = "Physical attributes (O4)"
 "P1" = "Education in general (P1)"
 "Q1" = "Linguistic Actions, States And Processes; Communication (Q1)"
 "Q2" = "Speech (Q2)"
 "Q3" = "Language, speech and grammar (Q3)"
 "Q4" = "The Media (Q4)"
 "S1" = "Social Actions, States And Processes (S1)"
 "S2" = "People (S2)"
 "S3" = "Relationship (S3)"
 "S4" = "Kin (S4)"
 "S5" = "Groups and affiliation (S5)"
 "S6" = "Obligation and necessity (S6)"
 "S7" = "Power relationship (S7)"

```

"S8" = "Helping/hindering (S8)"
"S9" = "Religion and the supernatural (S9)"
"T1" = "Time (T1)"
"T2" = "Time: Beginning and ending (T2)"
"T3" = "Time: Old, new and young; age (T3)"
"T4" = "Time: Early/late (T4)"
"W1" = "The universe (W1)"
"W2" = "Light (W2)"
"W3" = "Geographical terms (W3)"
"W4" = "Weather (W4)"
"W5" = "Green issues (W5)"
"X1" = "Psychological Actions, States And Processes (X1)"
"X2" = "Mental actions and processes (X2)"
"X3" = "Sensory (X3)"
"X4" = "Mental object (X4)"
"X5" = "Attention (X5)"
"X6" = "Deciding (X6)"
"X7" = "Wanting; planning; choosing (X7)"
"X8" = "Trying (X8)"
"X9" = "Ability (X9)"
"Y1" = "Science and technology in general (Y1)"
"Y2" = "Information technology and computing (Y2)"
"Z0" = "Unmatched proper noun (Z0)"
"Z1" = "Personal names (Z1)"
"Z2" = "Geographical names (Z2)"
"Z3" = "Other proper names (Z3)"
"Z4" = "Discourse Bin (Z4)"
"Z5" = "Grammatical bin (Z5)"
"Z6" = "Negative (Z6)"
"Z7" = "If (Z7)"
"Z8" = "Pronouns (Z8)"
"Z9" = "Trash can (Z9)"
"tokens" = "Text length"
;
run;
quit;
ods html file="&whereisit/&myfolder/loadtable.html";
%macro create(howmany);
%do i=1 %to &howmany;

title "LOADINGS TABLE";
title2 "Factor &i pos" ;
data temp;
set rotated4 ;
where factor="fac&i" and pole=1 ;
proc sort;
by descending Factor&i ;
proc print ; FORMAT _NAME_ $semfeaturelabels.; var _NAME_ Factor&i ;
run;

title "Factor &i neg" ;
data temp;
set rotated4 ;
where factor="fac&i" and pole=-1 ;

```

```

proc sort;
  by Factor&i ;
proc print ; FORMAT _NAME_ $semfeaturelabels.; var _NAME_ Factor&i ;
run;

%end;
%mend create;
%create(&extractfactors)
ods html close;
quit;

PROC EXPORT
  DATA= user.rotated3
  DBMS=TAB
  OUTFILE="&whereisit/&myfolder/rotated.tsv"
  REPLACE;
RUN;

/*END PART 15*/
/* BEGINNING PART 16*/

/* factor scores */
/* no standardizing the data because it is ranked */

data rotated4; set rotated3; if loaded = 1; run;

proc sort data=rotated4;
  by factor ;
run;
proc transpose data=rotated4 out=score;
  by factor ;
  id _NAME_ ;
  var pole;
run;

data score;
  _type_='SCORE';
  set score;
  drop _name_;
  rename factor=_name_;
run;
proc score data=ranked score=score out=scores; run;

/* turn missing values to zeros */

proc stdize data = scores out=scores reponly missing=0;
var &highcomm ;
run;

data temp (keep= file tokens) ; set typetoken;
proc sort; by file;
run;

proc sort data = scores ; by file ; run;

```

```

data scores;
merge scores (in=a) temp (in=b) ;
by file;
if (a and b) then output;
run;

data scores_only
(keep = file &factorvars ) ;
set scores ;
run;

PROC EXPORT
DATA= user.scores
DBMS=TAB
OUTFILE="&whereisit/&myfolder/&project._scores.tsv"
REPLACE;
RUN;

PROC EXPORT
DATA= user.scores_only
DBMS=TAB
OUTFILE="&whereisit/&myfolder/&project._scores_only.tsv"
REPLACE;
RUN;

proc corr data=scores; var tokens; with &factorvars; run;

data temp (keep= file &factorvars tokens);
set scores;
proc sort; by descending fac1;
run;

proc print data=temp (FIRSTOBS=1 OBS=20) ; run;

data temp (keep= file &factorvars tokens);
set scores;
proc sort; by fac1;
run;

proc print data=temp (FIRSTOBS=1 OBS=20) ; run;

```

ANEXO 3 - PROGRAMA SAS PARA ANÁLISE FATORIAL ACÚSTICA

```

/* BEGINNING PART 1 */
/* ==== EDIT BELOW =====*/

```

```

%let project = claudia_acoustic ;

%let myfolder = &project ;

%let sasusername = u50418389 ;

%let whereisit = /home/&sasusername ; /* online */
libname user "&whereisit/&myfolder";

options fmtsearch=(user work library);

/* enter number of factors to extract */
%let extractfactors = 3 ;

/* enter factor variable names */
%let factorvars = fac1-fac3 ;

/* enter min loading cutoff */
%let minloading = .3 ;

/* enter min communality cutoff */
%let communalcutoff = .15 ;

/* END PART 1 */
/* BEGINNING PART 2 */

DATA observed (DROP= line1-line16) ;
INFILE "&whereisit/&myfolder/&project._counts.txt";
input line1 folder
#2 line2 file
#3 line3 danceability
#4 line4 energy
#5 line5 key
#6 line6 loudness
#7 line7 mode
#8 line8 speechiness
#9 line9 acousticness
#10 line10 instrumentalness
#11 line11 liveness
#12 line12 valence
#13 line13 tempo
#14 line14 id $ 4-25
#15 line15 duration
#16 line16 timesgn
;
RUN;

/* delete files that don't match the lyrics files */

DATA deletfiles ;
INFILE "&whereisit/&myfolder/files_to_delete_claudia_annotated_2.txt";
input filed
;

```

```

RUN;

proc sql;
  create table temp as select * from observed
  where file not in(select filed from deletefiles );
quit;

proc sort data=temp;
by file;
run;

data observed ;
set temp;
run;

/* delete suspect tracks based on speechiness and instrumentalness */

data temp1;
set observed;
IF speechiness <= .71 ;
run;

data temp2;
set temp1;
IF speechiness > 0 ;
run;

data temp3;
set temp2;
IF instrumentalness <= .6 ;
run;

PROC SORT data=temp3;
by descending instrumentalness ;
run;

PROC EXPORT
  DATA= user.observed
  DBMS=TAB
  OUTFILE="&whereisit/&myfolder/observed.tsv"
  REPLACE;
RUN;

/*BEGINNING PART */

/* unrotated, before dropping low communalities */

proc datasets library=user nolist;
delete
fout;
run;

ODS EXCLUDE NONE;

```

```

proc factor fuzz=0.3 data= observed OUTSTAT= fout NOPRINT
method=principal
plots=scree
mineigen=1
reorder
heywood
nfactors=100
;
var danceability -- tempo timesgn ;
run;

/*** find low communalities ***/

data fout2;
  set fout (where=( _TYPE_="COMMUNAL"));
run;

proc transpose data=fout2 out=communal; id _TYPE_; run;
/* list vars to drop */
proc sql ;
  select _name_ into :lowcomm separated by ' ' from communal
  where communal < &communalcutoff ;
quit;

/* list vars to keep */

ODS EXCLUDE NONE ;
proc sql ;
  select _name_ into :highcomm separated by ' ' from communal
  where communal >= &communalcutoff ;
quit;

/* save communalities to spreadsheet */

PROC SORT data=communal (keep= _name_ communal); BY communal ; RUN;

PROC EXPORT
DATA= USER.communal
DBMS=TAB
OUTFILE="&whereisit/&myfolder/communalities.tsv"
REPLACE;
RUN;

/* scree plot */

data fout2;
  set fout (where=( _TYPE_="EIGENVAL"));
run;

proc transpose data=fout2 out= fout3 (drop = _NAME_);
id _TYPE_;
run;

data fout4 ;

```

```

set fout3 ;
factor = _n_;
if factor <= 20 ;
run;

/* delete the scree files */
FILENAME temp "/home/&sasusername/&myfolder/scree_1.png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

FILENAME temp "/home/&sasusername/&myfolder/scree_2.png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

%macro create(howmany);
%do i=1 %to &howmany;

FILENAME temp "/home/&sasusername/&myfolder/scree_1&i..png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;
FILENAME temp "/home/&sasusername/&myfolder/scree_2&i..png";
DATA _NULL_;
rc=FDELETE('temp');
RUN;

%end;
%mend create;
%create(9)
;

/* create the scree files */

ods listing gpath="&whereisit/&myfolder/";
ods graphics / imagename="scree_1" imagefmt=png;
title "Scree plot";
proc sgplot data= fout4 ;
series x=factor y=EIGENVAL / datalabel=EIGENVAL;
run;
title;

ods listing gpath="&whereisit/&myfolder/";
ods graphics / imagename="scree_2" imagefmt=png;
title "Scree plot";
proc sgplot data= fout4 ;
series x=factor y=EIGENVAL / datalabel=factor;
run;
title;

/* determine the number of factors and enter number in the header of this program */

```

```

/* END PART 12 */
/* BEGINNING PART 13 */

/* rotated w/o low communalities */
/* ignore error 'matrix is singular' because results will be output anyway */

proc datasets library=user nolist;
delete
rotatedfinal ;
run;

ods html file="&whereisit/&myfolder/rotated.html";
proc factor fuzz=0.1 data= observed OUTSTAT= rotatedfinal
method=principal
msa
mineigen=0
priors=max /* max = matrix is singular */
nfactors= &extractfactors
rotate=promax
heywood
;
var &highcomm ;
run;
ods html close;

/* factor scores */

PROC STANDARD DATA=observed MEAN=0 STD=1 OUT=mdz;
var
energy loudness tempo acousticness danceability
valence speechiness liveness instrumentalness mode key ;
RUN;

DATA scores;
set mdz ;
fac1 = ( energy + loudness + liveness + tempo ) - acousticness;
fac2 = ( danceability + valence + timesgn + speechiness ) - instrumentalness ;
fac3 = mode - key ;
run;

data temp4 (KEEP= file fac1 fac2 fac3 ) ;
set scores;
proc sort; by file;
run;

data temp5 (DROP= folder );
set observed ;
proc sort; by file;
run;

data scores;
merge temp4 temp5 ;
by file;

```

```

run;

PROC EXPORT
DATA= USER.scores
DBMS=TAB
OUTFILE="&whereisit/&myfolder/scores.tsv"
REPLACE;
RUN;

/* sample the data */
/* mean scores for top 1000 scores on each dim */

/* positive */

%macro create(howmany);
%do i=1 %to &howmany;
proc sort data=scores; by descending fac&i ;
run;
data samplef&i.pos;
set scores;
if _n_ <=1000 ;
run;

ods html file="&whereisit/&myfolder/means_samplef&i.pos.html";
proc means ; run;
ods html close;

%end;
%mend create;
%create( &extractfactors ) /* number of factors extracted */
ods html close;
quit;
/* negative pole */
%macro create(howmany);
%do i=1 %to &howmany;
proc sort data=scores; by fac&i ;
run;
data samplef&i.pos;
set scores;
if _n_ <=1000 ;
run;

ods html file="&whereisit/&myfolder/means_samplef&i.neg.html";
proc means ; run;
ods html close;

%end;
%mend create;
%create( &extractfactors ) /* number of factors extracted */
ods html close;
quit;

```

ANEXO 4 - PROGRAMA SAS PARA ANÁLISE CANÔNICA

```
/* BEGINNING PART 1 */
/* ==== EDIT BELOW =====*/

/* enter type of MD analysis */

%let project = claudia_canonical ;

%let myfolder = &project ;

%let sasusername = u50418389 ;

%let whereisit = /home/&sasusername ;

libname user "&whereisit/&myfolder";

options fmtsearch=(user work library);

/* END PART 1 */
/* BEGINNING PART 2 */

/* read in acoustic data */

FILENAME IN "&whereisit/claudia_acoustic/scores.tsv";
PROC IMPORT OUT= acoustic
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

/* read in semantic data */
FILENAME IN "&whereisit/claudia_semantic/claudia_semantic_scores_only.tsv";
PROC IMPORT OUT= semantic
  DATAFILE= IN
  DBMS=DLM REPLACE;
  GETNAMES=YES;
  delimiter='09'x;
  datarow=2;
  GUESSINGROWS=1000;
RUN;

data acoustic (KEEP= file fac1 fac2 fac3
  RENAME = (fac1=afac1 fac2=afac2 fac3=afac3));
set acoustic;
run;

data semantic (RENAME = (fac1=sfac1 fac2=sfac2 fac3=sfac3
  fac4=sfac4 fac5=sfac5));
```

```

set semantic;
run;

/* merge the two datasets */

data merged;
merge acoustic (in=a) semantic (in=b) ;
by file;
if (a and b) then output;
run;

PROC EXPORT
DATA= user.merged
DBMS=TAB
OUTFILE="&whereisit/&myfolder/merged.tsv"
REPLACE;
RUN;

ods html file="&whereisit/&myfolder/canonical.html";
proc cancorr corr data=merged;
var sfac1-sfac5 ;
with afac1-afac3 ;
run;
ods html close;

ods html file="&whereisit/&myfolder/canonical2.html";
proc cancorr data=merged out=canout
vprefix=semantic vname="Sem. dims."
wprefix=musical wname="Mus. dims." ;
var sfac1-sfac5 ;
with afac1-afac3 ;
run;
ods html close;

PROC EXPORT
DATA= user.canout
DBMS=TAB
OUTFILE="&whereisit/&myfolder/canonical.tsv"
REPLACE;
RUN;

proc gplot;
axis1 length=5.5 in;
axis2 length=5.5 in;
plot semantic2*musical2 / vaxis=axis1 haxis=axis2;
symbol v=J f=special h=2 i=r color=blue;
run;

```