

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

**Patrícia Bértoli-Dutra**

***LINGUAGEM DA MÚSICA POPULAR ANGLO-AMERICANA  
DE 1940 A 2009***

**DOUTORADO EM LINGUÍSTICA APLICADA E  
ESTUDOS DA LINGUAGEM**

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

**SÃO PAULO**

**2010**

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

BANCA EXAMINADORA

---

---

---

---

---

## Ficha Catalográfica

**BÉRTOLI-DUTRA**, Patrícia, *Linguagem da música popular anglo-americana de 1940 a 2009*. São Paulo: 2009, pp.270.

**Tese (Doutorado)** – Pontifícia Universidade Católica de São Paulo, 2010

**Área de concentração:** Linguística Aplicada e Estudos da Linguagem

**Orientador:** Professor Doutor Antonio Paulo Berber Sardinha

**Palavras-chave:** letras de música; dimensões; variação linguística

## AUTORIZAÇÃO

Na qualidade de autora, autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução parcial ou total desta dissertação por processos fotocopiadores ou eletrônicos

São Paulo, 31 de março de 2010.

**Para minha mãe e meu pai.**

**Para minhas irmãs.**

**Pois juntos continuamos um perfeito AMAPA**

## AGRADECIMENTOS

Agradeço, primeiramente a Deus, que é a luz que guia meu caminho enquanto me protege, que me permitiu chegar até aqui para que eu possa compartilhar tudo o que aprendi.

Em seguida à CAPES, por ter fornecido apoio financeiro durante o curso tanto no Brasil quanto durante o estágio no exterior, por meio do auxílio bolsa-sanduiche.

Ao meu orientador Tony, que ao me apresentar à Linguística de Corpus na época do mestrado fez-me apaixonar pelos valores numéricos das palavras. Agradeço ter compartilhado seu imenso conhecimento comigo, com paciência, ternura, entusiasmo e, sobretudo, acreditando em mim incondicionalmente.

Aos professores que fizeram parte de minhas bancas de qualificação e de defesa, cujos conselhos colaboraram para direcionar minha pesquisa para o caminho que ela assumiu hoje: Profa. Dra. Heloísa Collins; Profa. Dra. Tânia Shepphard (pela generosidade, bondade e sapiência); Profa. Dra. Ana Rachel Machado (pelo melhor “jeito estúpido de ser” que alguém pode ter); Profa. Dra. Deise Prina Dutra (seus conselhos superam qualquer expectativa); Prof. Dr. Celso Mojola (pelo valor e respeito pela música); Prof. Flávio Barbosa e Prof. Marcello Marcelino.

Ao Prof. Dr. Scott Crossley, por seu apoio, estímulo e orientação durante minha visita à Mississippi State University, EUA, e a todos do departamento de inglês da MSU que gentilmente me acolheram. Aos amigos que fiz nos EUA, brasileiros e estrangeiros, extremamente estudiosos e dedicados à pesquisa.

Ao grupo de amigos orientandos do Tony, cujas contribuições foram inestimáveis para o desenvolvimento do trabalho e para a análise dos dados e cuja amizade possibilitou o andamento do curso de maneira descontraída e bem direcionada, sobretudo ao Carlos Kauffmann pela gentileza com que compartilha seu conhecimento.

À Ciça Lopes, sem a qual eu, definitivamente, não teria terminado, por seu envolvimento nas questões burocráticas do meio acadêmico (representando-me nas matrículas, pelas escadas subidas, etc.), por compartilhar discussões, realizações e preocupações, revisões, enfim, pela amizade constante.

À Cristiane Magalhães, pelo companheirismo, pela amizade, pelo incentivo, pela irmandade, por enxergar em mim aquilo que eu mesma não vejo, pela admiração recíproca.

Aos funcionários do LAEL, especialmente à Maria Lúcia, muito mais do que uma secretária, mas uma amiga paciente, preocupada e motivadora.

A todos os meus alunos, que são o motivo desta pesquisadora estar sempre em busca de opções para o ensino, em especial ao Bruno pela colaboração na coleta do corpus.

Ao meu amigo prof. Dr. Tito Damazo, pelo modelo que é para mim, pelo apoio e cumplicidade de sempre, sobretudo pela revisão incomparável.

A Adriana Segura, minha amiga, que sempre ‘resolve’ a minha vida.

Aos colegas do grupo de inglês que não cessam de procurar “um inglês melhor”, em especial ao Arthur pela constante inspiração e motivação.

Ao meu marido Otávio, pelo companheirismo inigualável e por respeitar minha necessidade de solidão.

A meu filho, por ser o filho perfeito (para mim), solidário e atencioso e por colocar duas novas fontes de alegria em minha vida: Laiz e Lara.

Por me proporcionarem a paz quando eu mais precisava, agradeço às minhas tias Clarissinha, Eliana, Elzita e Zoraide e ao meu amado tio Paulo.

Existe em mim a necessidade de oferecer um agradecimento especial, que vai além da dedicatória, às minhas irmãs Andréa e Anelise por seu apoio. Vocês são a fonte da força que me sustenta e é por isso que eu consigo fazer o que faço.

Finalmente, ainda sobre a dedicatória, agradeço especialmente a minha mãe e a meu pai, a quem devo tudo o que tenho e o que sou. Por me ensinarem que ser uma pessoa do bem leva à riqueza da alma e à paz de espírito, que tudo aquilo que não é compartilhado, especialmente o conhecimento, é perdido. Por terem conquistado e ensinado que a conquista se dá através de estudo, trabalho, honestidade, educação e, sobretudo, fé.

## RESUMO

O objetivo desta pesquisa é desvendar as dimensões de variação linguística que permitem uma caracterização da música popular anglo-americana a partir da observação de letras de música. Para tanto a pesquisa encontrou suporte teórico na Linguística de Corpus (BIBER, 1988; SINCLAIR, 1991; BERBER SARDINHA, 2004), que parte da perspectiva probabilística e funcional da língua, assumindo que as variações linguísticas ocorrem de acordo com seu contexto (FIRTH, 1957; HALLIDAY; HASAN, 1989; HALLIDAY, 1991; HALLIDAY; WEBSTER; 2002). A música popular foi considerada por sua representatividade social (MOORE, 2003; STARR; WATERMAN, 2007) e sua representação textual, as letras, observadas enquanto fonte de investigação linguística. Assim, adotamos a metodologia de análise proposta por Biber (1988), a Análise Multidimensional, que prevê a investigação de textos não apenas a partir de uma característica, mas sob diversas perspectivas representativas de determinadas funções linguísticas. Coletamos um corpus contendo aproximadamente 1.200.000 palavras, advindas de 6.290 letras de música compostas originalmente em inglês. O corpus foi etiquetado em relação às suas características morfossintáticas e semânticas. Essas características e os pacotes lexicais de três palavras (trigramas) mais frequentes no corpus e no inglês geral (corpus Google N-Gram) foram considerados como variáveis de análise e fornecidos ao programa de extração fatorial SPSS. A análise fatorial reduz o enorme número de variáveis, agrupando-as conforme sua co-ocorrência, por meio da identificação das variáveis que se distribuem de forma semelhante. As 97 variáveis iniciais da pesquisa foram agrupadas em 13 variáveis gramaticais, 8 semânticas e 2 de padronização (trigramas). A análise fatorial apresentou três fatores para cada grupo que foram interpretados em sete dimensões diferentes: argumentação versus informação; argumentação e padronização; interação versus descrição; narrativa no passado versus referencial imediato; ações pessoais; emoção e sociedade e referência a música. A análise de letras de música nos eixos dimensionais apresentou o modo como artistas, gêneros musicais e os períodos de tempo se distribuem linguisticamente.



## ABSTRACT

This research aims at finding the dimensions of linguistic variation in order to categorize Anglo-American popular music, by observing their lyrics. The main theoretical framework for the research is provided by Corpus Linguistics (BIBER, 1988; SINCLAIR, 1991; BERBER SARDINHA, 2004), which views the language from a probabilistic and functional perspective, assuming that linguistic variation occurs according to the context (FIRTH, 1957; HALLIDAY; HASAN, 1989; HALLIDAY, 1991; HALLIDAY; WEBSTER; 2002). Popular music was considered for its social relevance (MOORE, 2003; STARR; WATERMAN, 2007) and its textual representation, the lyrics, considered as a source for linguistic investigation. Thus, we selected Biber's model for a Multi-dimension analysis, which predicts that texts should be analyzed not only taking into account one but several linguistic features so as to determine their variation across linguistic functions. The corpus collected for the study is consisted of approximately 1,200,000 words from 6,290 song lyrics originally written in English. The corpus was tagged for its parts-of-speech features and for its semantic groupings. These features and the most frequent lexical bundles (3-grams) in the corpus and in general English (Google N-Gram corpus) were considered as variables for the factor extraction at the SPSS program. Factor analysis reduces the huge number of variables, grouping them according to their co-occurrence. This procedure is done through the identification of the distribution patterns of variables. The 97 initial variables in our research were grouped into 13 grammar variables, 8 semantic variables, and 2 pattern variables (3-grams). Factor analysis resulted in three factors for each of those groups. From their interpretation seven different dimensions emerged: argumentative versus informative; argumentative and pattern; interactive versus descriptive; past narratives versus immediate context; personal acts; emotion and society; and musical manifestation. The analysis of song lyrics on the dimensional scale showed how singers and bands, musical styles and the decade of the recordings are closer or more distant to each other in linguistic terms.

# SUMÁRIO

<b>INTRODUÇÃO</b>	<b>1</b>
1. Motivação e trajetória da pesquisa	1
2. Pressupostos e conceitos	2
3. Objetivos e questões de pesquisa	6
4. Áreas da pesquisa	8
5. Organização da tese	10
<b>CAPÍTULO 1: Música e identidade cultural, sociedade e língua</b>	<b>12</b>
1.1. Música e identidade	12
1.2. Música e Vídeo	16
1.3. Música e armazenagem	18
1.4. Música e linguagem	19
1.5. Música Popular	24
1.6. Trajetória da Música Anglo-Americana	27
<b>CAPÍTULO 2: FUNDAMENTAÇÃO TEÓRICA</b>	<b>41</b>
2.1. Linguística de Corpus	41
2.2. Análise Multidimensional	50
2.3. Gênero e letras de música	70
<b>CAPÍTULO 3: ESTUDO PILOTO</b>	<b>76</b>
3.1. Metodologia	77
3.1.1. Corpus de estudo	77
3.1.2. Corpora de referência	79
3.1.3. Procedimentos de Análise	81
3.2. Resultados	85
3.3. Conclusão	90
<b>CAPÍTULO 4: METODOLOGIA DE PESQUISA DO ESTUDO PRINCIPAL</b>	<b>92</b>
4.1. Os corpora utilizados	95
4.1.1. O corpus de estudo	95
4.1.2. Corpus Google N-Gram	100

4.2. Características Linguísticas e as variáveis da pesquisa	101
4.3. Extração de Fatores	112
4.3.1. Primeira Extração	119
4.3.2. Extração em três partes	124
4.3.2.1. Componentes gramaticais	124
4.3.2.2. Componentes semânticos	128
4.3.2.3. Componentes léxico-gramaticais	130
4.4. Fatores	133
<b>CAPÍTULO 5: RESULTADOS E DISCUSSÃO</b>	<b>136</b>
5.1. Interpretação dos fatores	137
5.1.1. Interpretando o Fator Gramatical Gram_C1	140
5.1.2. Interpretando o Fator Gramatical Gram_C2	144
5.1.3. Interpretando o Fator Gramatical Gram_C3	146
5.1.4. Interpretando o Fator Semântico Sem_C1	149
5.1.5. Interpretando o Fator Semântico Sem_C2	151
5.1.6. Interpretando o Fator Semântico Sem_C3	152
5.1.7. Interpretando os Fatores Léxico-gramaticais	155
5.2. Dimensões de variação na música popular anglo-americana	159
5.2.1. Dimensões Gramaticais	162
5.2.1.1. Primeira Dimensão Gramatical: Argumentação X Informação	163
5.2.1.2. Segunda Dimensão Gramatical: Interação X Descrição	176
5.2.1.3. Terceira Dimensão Gramatical: Narrativa no passado X referencial imediato	187
5.2.2. Dimensões Semânticas	196
5.2.2.1. Primeira Dimensão Semântica: Ações pessoais	198
5.2.2.2. Segunda Dimensão Semântica: Emoções e sociedade	206
5.2.2.3. Terceira Dimensão Semântica: Referência a música	214
5.2.3. Dimensões Léxico-Gramaticais	222
5.2.4. Síntese dos resultados	229
5.3. Discussão	230

<b>CAPÍTULO 6: CONCLUSÃO</b>	<b>233</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>242</b>
<b>ANEXOS</b>	<b>249</b>

#### **LISTA DE ANEXOS**

ANEXO 1 - LISTA DOS 200 TRIGRAMAS MAIS FREQUENTES DO CORPUS DE ESTUDO	250
ANEXO 2 - AMOSTRA DA PLANILHA DE ANÁLISE FATORIAL	251
ANEXO 3 - AMOSTRA DA PLANILHA DE ANÁLISE FATORIAL COM VALORES NORMALIZADOS E VARIÁVEIS REDUZIDAS	258
ANEXO 4 - LISTA DE PALAVRAS DE CLASSIFICAÇÃO SEMÂNTICA PARA A CATEGORIA ESTADO EMOCIONAL	261
ANEXO 5 - AMOSTRA DA PLANILHA DO EXCEL COM OS ESCORES PADRONIZADOS DOS COMPONENTES FATORIAIS PARA OS 30 PRIMEIROS ARQUIVOS DO NOSSO CORPUS DE ESTUDO	262
ANEXO 6 - CATEGORIZAÇÃO DOS 32 ARTISTAS DO NOSSO CORPUS DE ESTUDO POR ESTILO/GÊNERO MUSICAL	264
ANEXO 7 - IMAGEM DA ANÁLISE DESCRITIVA FINAL NO PROGRAMA SPSS PARA CADA COMPONENTE FATORIAL SOB A PERSPECTIVA “YEAR”	269
ANEXO 8 – QUADROS COM AS MÚSICAS MAIS FREQUENTES NA DIMENSÃO GRAMATICAL 1	270

#### **LISTA DE QUADROS**

Quadro 1: Classificação de corpora por número de palavras. Reproduzido de Berber Sardinha (2004a, p. 26)	43
Quadro 2: Resumo da estrutura fatorial obtida na pesquisa de Biber para os fatores 1, 2 e 3 (BIBER, 1988, p. 102)	59
Quadro 3: Resumo da estrutura fatorial obtida na pesquisa de Biber para os fatores 4, 5, 6 e 7 (BIBER, 1988, p. 103)	60
Quadro 4: Reprodução da tabela 1.1. Características definidoras de registro, gênero e estilo (BIBER, 2009, p. 16)	73
Quadro 5: descrição resumida da composição do corpus de estudo	78
Quadro 6: descrição detalhada do corpus de estudo	78
Quadro 7: composição do BNC por categoria	80
Quadro 8: 500 palavras mais frequentes no corpus do estudo piloto	82

Quadro 9: contraste de frequência das palavras mais frequentes no CE com o BNC e o ANC	86
Quadro 10: amostra de contraste de frequência de itens lexicais do CE com o BNC e o ANC	86
Quadro 11: 30 primeiros trigramas no CE contrastados com trigramas dos corpora de referência	89
Quadro 12: descrição do corpus principal CM (corpus de música)	95
Quadro 13: descrição detalhada do corpus de estudo (CM)	96
Quadro 14: Corpus de Música (CM) subdividido por década	98
Quadro 15: etiquetas morfossintáticas do <i>Penn-tree tagger</i>	102
Quadro 16: amostra de categorias semânticas (agente e dia) extraídas pelo programa <i>semantic tagger</i>	106
Quadro 17: 39 categorias semânticas incluídas em nossa análise	108
Quadro 18: primeiras linhas de nossa planilha com dados normalizados	111
Quadro 19: funções desempenhadas pelas variáveis gramaticais	121
Quadro 20: Quadro comparativo dos pesos das variáveis nos componentes gramaticais e nos componentes léxico-gramaticais	132
Quadro 21: amostra da tabela de resultados de escore padronizado dos componentes fatoriais	133
Quadro 22: reprodução da matriz fatorial para as características gramaticais consideradas	138
Quadro 23: reprodução da matriz fatorial para as características semânticas consideradas	138
Quadro 24: reprodução da matriz fatorial para as características léxico-gramaticais consideradas	139
Quadro 25: oposição de polaridade entre os componentes do fator Lexgram_C1	157
Quadro 26: resumo da categorização dos estilos musicais por artista, em nosso corpus de estudo	161
Quadro 27: Estatística descritiva para o fator Gram_C1, em relação a cantor/banda	164
Quadro 28: Estatística descritiva para o fator Gram_C1, em relação ao ano	168
Quadro 29: Estatística descritiva para o fator Gram_C1, em relação estilo musical	171
Quadro 30: Estatística descritiva para o fator Gram_C2, em relação ao artista	178
Quadro 31: Estatística descritiva para o fator Gram_C2, em relação ao tempo	181
Quadro 32: Estatística descritiva para o fator Gram_C2, em relação ao gênero musical	183
Quadro 33: Estatística descritiva para o fator Gram_C3, em relação a artistas	188
Quadro 34: Estatística descritiva para o fator Gram_C3, em relação ao período de gravação	189
Quadro 35: Estatística descritiva para o fator Gram_C3, em relação ao estilo musical	189
Quadro 36: Resumo das dimensões gramaticais	196

Quadro 37: Estatística descritiva para o fator Sem_C1, em relação a artistas	198
Quadro 38: Estatística descritiva para o fator Sem_C1, em relação ao período de gravação	201
Quadro 39: Estatística descritiva para o fator Sem_C1, em relação ao estilo/gênero musical	203
Quadro 40: Estatística descritiva para o fator Sem_C2, em relação a artistas	207
Quadro 41: Estatística descritiva para o fator Sem_C2, em relação ao período de gravação	209
Quadro 42: Estatística descritiva para o fator Sem_C2, em relação ao gênero/estilo musical	210
Quadro 43: Estatística descritiva para o fator Sem_C3, em relação a artistas	214
Quadro 44: Estatística descritiva para o fator Sem_C3, em relação ao período de gravação	216
Quadro 45: Estatística descritiva para o fator Sem_C3, em relação ao estilo/gênero musical	217
Quadro 46: Resumo das dimensões semânticas	221
Quadro 47: Estatística descritiva para o fator LexGram_C1, em relação ao artista	223
Quadro 48: Estatística descritiva para o fator LexGram_C1, em relação ao período de gravação	224
Quadro 49: Estatística descritiva para o fator LexGram_C1, em relação ao estilo/gênero	225
Quadro 50: Resumo das dimensões léxico-gramaticais	229
Quadro 51: Representatividade nas dimensões de análise	238

## LISTA DE FIGURAS

Figura 1: CM, na década de 1960, com as músicas gravadas por Elvis Presley em 1966	99
Figura 2: imagem da planilha do Excel com as 32 primeiras linhas de nossa planilha, contendo as 15 primeiras variáveis (tokens; types; CC; CD ; DT; EX; FW; IN; JJ; JJS; LS; MD; NN; NNS)	110
Figura 3: imagem da primeira janela do programa de análise fatorial SPSS, para a seleção de variáveis	113
Figura 4: Variáveis selecionadas para a análise fatorial	114
Figura 5: imagem da janela de extração de análise fatorial, com suas seleções padrão ( <i>default</i> )	115
Figura 6: imagem de um Gráfico <i>scree</i> com todas as variáveis de nossa pesquisa	117
Figura 7: imagem da matrix fatorial resultante de nossa primeira extração	119
Figura 8: amostra da planilha para análise fatorial com variáveis agrupadas e dados normalizados	123
Figura 9: Quadro de análise fatorial com variáveis gramaticais selecionadas	125
Figura 10: Gráfico <i>scree</i> produto da extração dos componentes gramaticais	125

Figura 11: matriz padrão da extração de componentes gramaticais com 13 variáveis, 3 fatores (pesos acima de 0,30)	126
Figura 12: Gráfico <i>scree</i> produto da extração dos componentes semânticos	128
Figura 13: matriz padrão da extração de componentes semânticos com 08 variáveis, 3 fatores (pesos acima de 0,30)	129
Figura 14: matriz padrão da extração de componentes léxico-gramaticais (12 componentes gramaticais mais as contagens de trigramas)	131
Figura 15: Imagem da tela de cálculo de variáveis “ <i>compute variable</i> ” do programa SPSS para as variáveis do fator 1, da instância gramatical	134
Figura 16: Imagem da tabela do Excel com o valor de cada fator calculado para cada um dos textos do corpus	140
Figura 17: tabela do Excel com as 10 primeiras linhas de nosso arquivo acrescidas das classificações por ano, artista e gênero.	162

## LISTA DE TEXTOS

Texto 1: Letra da música <i>Should I</i> – gravação de Frank Sinatra, 1950	141
Texto 2: Letra da música <i>Candy perfume girl</i> – gravação de Madonna, 1998	141
Texto 3: Letra da música <i>Yours, sincerely</i> – gravação de Ray Charles, 2003	143
Texto 4: Letra da música <i>Forever</i> – gravação dos Beach Boys, 1970	144
Texto 5: Letra da música <i>Straight cold player</i> – gravação de Lenny Kravitz, 1998	145
Texto 6: Letra da música <i>If I had three wishes</i> – gravação de Frank Sinatra, 1955	146
Texto 7: Letra da música <i>Whatever happened to Christmas</i> – gravação de Frank Sinatra, 1968	147
Texto 8: Letra da música <i>Because</i> – gravação dos Beatles, 1969	147
Texto 9: Letra da música <i>Green is the color</i> – gravação de Pink Floyd, 1969	148
Texto 10: Letra da música <i>At my window</i> – gravação dos Beach Boys, 1970	149
Texto 11: Letra da música <i>Mountain Greenery</i> – gravação de Frank Sinatra, 1963	150
Texto 12: Letra da música <i>Neiani</i> – gravação de Frank Sinatra, 1940	150
Texto 13: Letra da música <i>You never give me your money</i> – gravação dos Beatles, 1969	151
Texto 14: Letra da música <i>Burning man</i> – gravação de Queensryche, 1999	152
Texto 15: Letra da música <i>Rock’n’Roll music</i> – gravação de Beach Boys, 1976	153
Texto 16: Letra da música <i>My Bonnie</i> – gravação de Beatles, 1976	154
Texto 17: Letra da música <i>Guilty by association</i> – gravação de Madonna, 1994	156
Texto 18: Letra da música <i>My number one</i> – gravação de Paramore, 2009	157
Texto 19: Letra da música <i>Grow up</i> – gravação de Simple Plan, 2002	165
Texto 20: Letra da música <i>Shame on you</i> – gravação de Aerosmith, 1985	167
Texto 21: Letra da música <i>You will soar</i> – gravação de Creed, 2005	169

Texto 22: Letra da música <i>Radio Waves</i> – gravação de Pink Floyd/Roger Waters, 1987	170
Texto 23: Letra da música <i>Swim in Silence</i> – gravação de Paramore, 2009	171
Texto 24: Letra da música <i>Pom pom play girl</i> – gravação dos Beach Boys, 1964	172
Texto 25: Letra da música <i>Justify my love</i> – gravação de Madonna, 2000	179
Texto 26: Letra da música <i>Yet another movie</i> – gravação de Pink Floyd, 1987	180
Texto 27: Letra da música <i>Oh Bess</i> – gravação de Frank Sinatra, 1940	181
Texto 28: Letra da música <i>Nobody Wins</i> – gravação de Frank Sinatra, 1973	182
Texto 29: Letra da música <i>It's your baby, you rock it</i> – gravação de Elvis Presley, 1970	183
Texto 30: Letra da música <i>On the turning away</i> – gravação de Pink Floyd, 1987	184
Texto 31: Letras das músicas <i>You beat all I ever saw</i> e <i>I saw a man</i> – gravação de Johnny Cash, 1969 e 1973	190
Texto 32: Letra da música <i>Hit the lights</i> – gravação de Metallica, 1983	191
Texto 33: Letra da música <i>Is that all?</i> – gravação do U2, 1980	192
Texto 34: Letra da música <i>Die with your boots on</i> – gravação do Iron Maiden, 1983	192
Texto 35: Letra da música <i>Morning Christmas</i> – gravação de Beach Boys, 1964	199
Texto 36: Letra da música <i>She's too tough</i> – gravação de Def Leppard, 1993	200
Texto 37: Letra da música <i>Emily</i> – gravação de Frank Sinatra, 1963	202
Texto 38: Letra da música <i>Bambi Slaughter</i> – gravação do Nirvana, 2004	203
Texto 39: Letra da música <i>Harvester of sorrow</i> – gravação do Metallica, 1988	207
Texto 40: Letra da música <i>I believe</i> – gravação de Frank Sinatra, 1943	209
Texto 41: Letra da música <i>Moonchild</i> – gravação de Iron Maiden, 1988	210
Texto 42: Letra da música <i>I write the songs</i> – gravação de Frank Sinatra, 1994	215
Texto 43: Letra da música <i>If loveliness were music</i> – gravação de Frank Sinatra, 1946	216
Texto 44: Letra da música <i>Restless Heart Syndrome</i> – gravação do Greenday, 2009	217
Texto 45: Letra da música <i>Rip it up</i> – gravação de Elvis Presley, 1956	220
Texto 46: Letra da música <i>Shortenin' Bread</i> – gravação dos Beach Boys, 1979	226
Texto 47: Letra da música <i>Memory Remains</i> – gravação dos Metallica, 1997	227

## LISTA DE GRÁFICOS

Gráfico 1: Distribuição de estilos musicais na dimensão gramatical 1: Argumentação X Informação	175
Gráfico 2: Distribuição de período de gravação na dimensão gramatical 1: Argumentação X Informação	175
Gráfico 3: Distribuição de artistas na dimensão gramatical 1:	



Argumentação X Informação	176
Gráfico 4: Distribuição de artistas na dimensão gramatical 2: Interação X Descrição	185
Gráfico 5: Distribuição de período de gravação na dimensão gramatical 2: Interação X Descrição	186
Gráfico 6: Distribuição de estilos musicais na dimensão gramatical 2: Interação X Descrição	186
Gráfico 7: Distribuição de artistas na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato	194
Gráfico 8: Distribuição de período de gravação na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato	195
Gráfico 9: Distribuição de estilos musicais na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato	195
Gráfico 10: Distribuição de artistas na dimensão semântica 1: Ações Pessoais	204
Gráfico 11: Distribuição por período de gravação na dimensão semântica 1: Ações Pessoais	205
Gráfico 12: Distribuição por período de gravação na dimensão semântica 1: Ações Pessoais	205
Gráfico 13: Distribuição de artistas na dimensão semântica 2: Emoções e Sociedade	212
Gráfico 14: Distribuição por estilo/gênero musical na dimensão semântica 2: Emoções e Sociedade	212
Gráfico 15: Distribuição por período de gravação na dimensão semântica 2: Emoções e Sociedade	213
Gráfico 16: Distribuição de artistas na dimensão semântica 3: referência a música	218
Gráfico 17: Distribuição de período de gravação na dimensão semântica 3: referência a música	219
Gráfico 18: Distribuição de período de gravação na dimensão semântica 3: referência a música	220

## INTRODUÇÃO

### 1. Motivação e Trajetória da pesquisa

Esta pesquisa apresenta o processo de investigação de características linguísticas em letras de música gravadas originalmente em inglês. Em outras palavras, analisamos letras de música em forma de texto, verificando como se relacionam os itens lexicais em termos morfológicos, sintáticos e semânticos, contrastados com as ocorrências desses mesmos itens no inglês geral, verificando a existência de diferentes dimensões de análise para esse tipo de texto.

A motivação inicial desta investigação está ligada a minha frequente inquietação com relação ao uso de letras de música em atividades de ensino de línguas. Se, por um lado, a professora de inglês que sou vivencia a constante utilização de letras de música na própria sala da aula e também nas aulas de colegas professores de língua estrangeira e de língua materna – assim como a presença desse tipo de atividades em livros didáticos e para-didáticos (o que ocorre desde a educação infantil) –, por outro, a pesquisadora se depara com a quase ausência de subsídios teóricos que confirmam às letras de música o valor que a prática pedagógica vem lhe conferindo.

Essa discussão teve início em minha pesquisa de mestrado (BÉRTOLI-DUTRA, 2001), na qual criei e validei atividades com letras de música, seguindo os preceitos da Linguística de Corpus, sugerindo extrapolar os limites dos tradicionais exercícios de “fill-in-the-blanks”<sup>1</sup> relegados ao final das aulas, associando-os ao inglês em uso (autêntico) por meio de tarefas relacionadas ao mundo real do aprendiz. O corpus coletado na época tornou-se a pedra fundamental da pesquisa que ora descrevemos e era composto por cerca de 350 mil palavras, advindas de letras de música originalmente gravadas em inglês.

Propusemo-nos<sup>2</sup> a dar continuidade ao processo de desenvolvimento de atividades de ensino com letras de música que estivessem relacionadas a teorias linguísticas e de ensino de segunda língua, quando percebemos a necessidade de

---

<sup>1</sup> Exercícios nos quais os alunos preenchem espaços em branco, que se referem a palavras ausentes nas letras das músicas, enquanto ouvem a gravação de áudio.

<sup>2</sup> De agora em diante passo a usar o pronome pessoal em primeira pessoa do plural por considerar este trabalho fruto de um processo colaborativo e das várias vozes em minha cabeça.

trabalhar a letra de música de uma maneira mais informada, ou seja, percebemos a ausência de um trabalho científico que descrevesse componentes linguísticos presentes em letras de música que pudessem ser aproveitados sistematicamente para o ensino de língua, mais especificamente de língua inglesa.

Passamos a investigar mais detalhadamente a proximidade lexical do discurso de letras de música com o discurso coloquial, a partir das evidências obtidas em nossa pesquisa de mestrado, desenvolvendo um estudo de convergência lexical entre a linguagem da música popular anglo-americana e a língua inglesa em geral. O corpus de estudo que utilizamos é resultado da constante construção do anterior, passando a contar, nessa fase, com cerca de um milhão de palavras. Esse estudo sobre convergência lexical constitui aquele que chamamos de estudo piloto, o qual será apresentado em detalhes no capítulo 3. Trata-se de um piloto pois embora tenhamos descoberto uma nova perspectiva de aproximação entre a língua inglesa geral e as letras de música, ele também apontou para a necessidade de considerarmos outros aspectos da língua, de ordem morfossintática e semântica, impulsionando a busca por uma nova metodologia de análise.

A proposta metodológica de Douglas Biber (BIBER, 1988) por uma análise multidimensional de traços linguísticos nos pareceu a mais adequada e nos permitiu o desenvolvimento do estudo principal que ora apresentamos. Todavia, preliminarmente ao detalhamento de nossos objetivos de pesquisa, apresentamos nossos pressupostos de pesquisa associados a alguns conceitos que permeiam o trabalho e as questões de pesquisa, juntamente com as áreas teóricas envolvidas, o que apresentaremos em seguida.

## **2. Pressupostos e conceitos**

Em nossa análise consideramos as características linguísticas das letras de música isoladas de seus componentes musicais, o que requer que expliquemos nossos pressupostos quanto a letras de música, música popular, fatores e características linguísticas.

Primeiramente, partimos da natureza sociológica da música popular, observando sua contribuição para a formação da cultura e da identidade dos indivíduos. A definição do gosto musical, associada a outros fatores sociais e culturais e as representações transferidas pela música – sejam elas exclusivamente sonoras ou a mensagem das letras ou o conteúdo imagético apresentado por seus intérpretes – passam a fazer parte do conjunto de características com as quais nos identificamos e, conseqüentemente construímos nossas personalidades, nossas identidades particulares e partilhadas.

No primeiro capítulo, apresentaremos uma discussão mais detalhada sobre música popular, no entanto é preciso evidenciar que a noção de música popular que permeia este trabalho é a mesma oferecida por Starr e Waterman (2007) que consideram como música popular a música de massa, ou seja aquela produzida e divulgada pela mídia a fim de atingir o maior número de consumidores possíveis. Adotamos essa perspectiva por considerarmos que esse tipo de música é aquele que chega rapidamente ao ouvido do consumidor, em especial do consumidor de música internacional – o que nos interessa de modo particular, pois prevemos o ensino de língua estrangeira.

Em outras palavras, partimos do pressuposto de que uma vez que músicas como as mencionadas acima têm maior probabilidade de fazer parte de nossas vidas, por seu componente midiático de massa, suas letras tornam-se fonte de conhecimento linguístico, permitindo a transferência das palavras que ouvimos nas letras para outras formas do discurso (e.g. quando conversamos sobre música, o que contribui para que confirmamos sentido àquilo que ouvimos) (MOORE, 2003). Griffiths (2003), por exemplo, comprova essa perspectiva ao confessar possuir um vídeo de uma apresentação da banda inglesa Oasis, em Knebworth, na Inglaterra, onde milhares de pessoas cantam todas as palavras de todas as músicas. Não se trata, contudo, de um fenômeno exclusivo de língua materna, pode ocorrer ainda que audiência e artistas sejam de nacionalidades e línguas diferentes. Basta ouvir as gravações de festivais como Rock' in Rio, por exemplo. Citamos, a título de ilustração, o caso da música *Love of my life*, da banda inglesa Queen, cuja gravação ao vivo feita em São Paulo, no ano de 1981, mostra que a voz do público se sobrepõe à voz do cantor, Freddie Mercury<sup>3</sup>.

---

<sup>3</sup> [http://www.youtube.com/watch?v=ToI\\_BqS4vuc](http://www.youtube.com/watch?v=ToI_BqS4vuc)

Essa visão popular da música nos remete a sua divisão em gêneros e estilos musicais populares com os quais diferentes ouvintes se identificam. Os rótulos também são impostos ao público pelos meios de consumo, que categorizam a música e seu público alvo, nem sempre exclusivamente por questões de afinidades sonoras. Veremos, ao longo deste trabalho, que a classificação em gêneros e estilos musicais proposta pelos meios midiáticos não é totalmente consensual entre especialistas e críticos da música popular anglo-americana. Optamos, portanto, por não fazer distinção entre os dois termos (gênero e estilo) neste trabalho. Ainda assim, entre os gêneros (ou estilos) que aparecem mais frequentemente na literatura de música popular encontram-se: *country* (*hillbilly*, *country and western*, *country pop*, *country alternativo*); jazz (*jazz funk*, *jazz rap*, *soul jazz*); *Rhythm and Blues* (*blues*, *soul*, *hip hop*, *rap*, *funk*); eletrônica (*dance*, *break*, *techno*); *pop* (música adulta, *disco*, *new romantic*, *new age*); *folk* (protesto, *new folk*, *folk alternativo*, cantor-compositor); *reggae*; *rock* (*rock'n'roll*, *hard rock*, *rock alternativo*, *emo*, *indie*, *folk rock*, *glam rock*, *punk*, *heavy metal – black*, *death*, *doom*, *metalcore*, *goth*, *thrash*) e *world music*. (SHUKER, 1994; BRACKETT, 2000; GOODWIN; GROSSBERG, 2003; MIDDLETON, 2003; MOORE, 2003; STARR; WATERMAN, 2007).

Num determinado momento de nossa pesquisa essa classificação genérica da música popular de massa se torna importante, pois o agrupamento das letras de música em termos lexicais será contrastado aos agrupamentos propostos por tal nomenclatura.

É preciso lembrar, contudo, que em nossa análise as letras de música são investigadas dissociadas de seus componentes sonoros. Essa singularização das letras de música implica em nosso pressuposto sobre letras de música enquanto texto integrante de um gênero, o que, conseqüentemente, implica no esclarecimento de nossa concepção sobre gênero.

Gênero, assim como registro, corresponde a uma diferente perspectiva de análise textual, sendo textos considerados como a forma, oral ou escrita, de linguagem natural utilizada para a comunicação. Nossa concepção segue Biber (1988; BIBER; CONRAD, 2009), que propõe que no nível do gênero são observados (e descritos) os propósitos comunicativos e o contexto situacional de textos, enfocando as estruturas de construção do texto como um todo. No nível do registro, são observadas as características

funcionais, representadas por características linguísticas (características sintáticas, morfológicas e semânticas) e como estas se associam aos propósitos comunicativos e situacionais do texto e/ou são motivadas por eles (BIBER; CONRAD, 2009, p. 1). Embora nossa pesquisa não seja sobre descrição de gêneros textuais ou discursivos, no capítulo teórico retomaremos essa discussão de forma breve.

Ao analisarmos letras de música como texto, representação da linguagem natural, estamos observando suas características no nível de registro e, por conseguinte, considerando seu pertencimento a um gênero, mais especificamente, ao gênero oral, tal como sermões e discursos políticos, os quais são preparados antecipadamente a serem apresentados ao ouvinte, mas não envolvem interação direta e imediata entre os interlocutores.

Segundo o sociólogo Frith (1998), as letras fazem parte do gênero coloquialismo conversacional, trazendo consigo inúmeras variações linguísticas, que permitem a exploração de diferentes línguas, modos de falar, de palavras enquanto símbolos sonoros, entre outros. Letras não são simplesmente poesias cantadas, mas uma forma de retórica e oratória, num relacionamento de persuasão entre cantor e ouvinte, que envolve a letra, o som e o movimento (a performance do intérprete). Podemos perceber, portanto, que a música pode ser associada a determinados contextos situacionais, carregando propósitos comunicativos; características próprias de um gênero.

Sendo assim, neste trabalho, pressupomos a existência do gênero letras de música, que se caracteriza principalmente por sua peculiaridade sonora, cujos textos compartilham o propósito de terem sido escritos de modo que pudessem ser integrados a estruturas melódicas. Os contextos de situação e propósitos comunicativos devem variar entre conjuntos de textos, podendo caracterizar subgêneros, o que não levaremos em conta para este trabalho especificamente. Todavia, no nível do registro serão investigadas características linguísticas desse gênero macro, observando variações linguísticas sob uma perspectiva multidimensional de análise.

Os termos características e traços linguísticos, variação e dimensão fazem parte da terminologia utilizada por Biber (BIBER, 1988; BIBER; CONRAD, 2009) na metodologia de análise que adotamos para este trabalho: Análise Multidimensional. *Características linguísticas*, também chamadas de *traços linguísticos* (BIBER, 1988;

BERBER SARDINHA, 2004a), representam os elementos lexicais e gramaticais típicos de um texto. Já *categorias linguísticas* são representadas pelas classes gramaticais (substantivos, adjetivos, conjunções, etc.). Biber observa que os traços linguísticos presentes nos textos não obedecem a um critério arbitrário, ou seja, existe variação nas escolhas. Em outras palavras, os falantes podem utilizar características linguísticas (palavras, estruturas, etc.) diferentes para dizer as mesmas coisas. A variação linguística pode ser associada às funções comunicativas do texto. Sendo assim, uma descrição de registro deve admitir vários níveis de análise, isto é diferentes dimensões.

Por se tratar de uma metodologia de cunho estatístico, que permite apreender a variação existente na língua, é necessário que cada característica lexical presente em cada texto seja considerada. Assim, os textos são etiquetados automaticamente por suas características linguísticas, ou seja, cada item lexical recebe uma etiqueta de acordo com a categoria linguística que representa. No programa computacional estatístico é desenvolvido um processo de análise fatorial, no qual cada característica linguística representa uma variável. Essas, por sua vez, são contrastadas umas com as outras em relação a cada um dos textos selecionados, a fim de se determinarem as correlações entre as variáveis, o que retornará um determinado número de fatores. Em outras palavras, os fatores representam os agrupamentos de variáveis que coocorrem. Cada fator corresponde, em parte, à variação linguística presente no conjunto de textos analisados. A partir da análise dos fatores é possível identificar as funções comunicativas, ou dimensões, subjacentes à língua (BIBER, 1988).

Feitos esses esclarecimentos iniciais quanto aos pressupostos de nossa pesquisa e pontuados conceitos e termos que permeiam nosso trabalho, apresentamos, a seguir, nossas perguntas e objetivos de pesquisa.

### **3. Objetivos e questões de pesquisa**

A presente pesquisa visa ao levantamento de dimensões de análise linguística para a música popular anglo-americana a partir de 97 variáveis (características

linguísticas), tais como, agrupamentos semânticos, características morfossintáticas e pacotes lexicais<sup>4</sup>, as quais estão explicitadas no capítulo metodológico.

Defendemos a tese de que a variação de estilos musicais, cantores e grupos, entre si e ao longo do tempo pode ser explicada em parte pela variação linguística das letras, não por meio de uma oposição binária, mas pela existência de dimensões de variação. Em outras palavras, se os rótulos impostos pela mídia para classificar a música popular de massa as distanciam – ou seja, o que é *rock* não é *punk*, que por sua vez não é *pop*, nem *country*, etc. –, as características linguísticas presentes nas letras de música devem demonstrar uma variação diferente, ou seja, devem existir letras específicas de *punk* que se aproximam das letras de música de *rock*, *pop* ou *country*. Isto é, segundo critérios linguísticos, muitas músicas *pop* podem ser mais ou menos próximas das de *rock*, enquanto muitas de *pop* podem ser bastante diferentes das de *rock* e assim por diante.

O objetivo principal da pesquisa descrita aqui é desvendar as dimensões de variação linguística que permitem uma caracterização da música popular anglo-americana a partir da observação de suas letras.

A fim de atingirmos os objetivos desta pesquisa, as seguintes perguntas orientaram nosso trabalho:

1. Quais os principais fatores<sup>5</sup> responsáveis pela variação linguística presente nas letras de música popular anglo-americana?
2. Quais dimensões (funções comunicativas) expressam os fatores obtidos?
3. Como essas dimensões são representadas em relação aos diferentes gêneros musicais presentes em nosso corpus de estudo?
4. Como essas dimensões são representadas em relação aos diferentes artistas selecionados para o estudo?

---

<sup>4</sup> Segundo Biber, Conrad e Cortes (2004, p. 376), pacotes lexicais, ou *lexical bundles*, são sequências de palavras que seguem padrões estruturais bastante constantes e que são mais recorrentes em um registro, ou seja, que tenha uma ocorrência mínima de 40 vezes em grupos de 1 milhão de palavras, considerando-se o mínimo de 5 textos distintos. Em nosso caso específico, consideramos agrupamentos de três palavras, também chamados de trigramas.

<sup>5</sup> Agrupamentos de variáveis (características linguísticas) coocorrentes, como dissemos anteriormente



5. Como essas dimensões são representadas ao longo do tempo por nós estudado (1940-2009)?

#### 4. Áreas da pesquisa

Nossa pesquisa está inserida na área de Linguística de Corpus<sup>6</sup>, área que se ocupa “da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” (BERBER SARDINHA, 2004a, p.3). Por considerar a língua um sistema probabilístico (HALLIDAY, 1991), a LC entende que as palavras ocorrem dentro de uma certa frequência, combinando-se e associando-se a outras palavras que definirão suas funções e seu uso.

A LC está muito ligada ao desenvolvimento tecnológico, pois o aumento da capacidade de armazenagem de dados cada vez maiores em menores espaços, associada ao desenvolvimento de programas (softwares) computacionais que são capazes de agrupar, categorizar, etiquetar gramatical e semanticamente, de modo automático, possibilitou a análise de grandes quantidades de texto partir da observação empírica dos dados.

É importante ressaltar que a LC não é uma área que possa ser considerada simplesmente como uma metodologia, ou um conjunto de ferramentas computacionais, pois, além das ferramentas, apresenta embasamento teórico e produção de “conhecimento novo”.

O maior representante da Linguística de Corpus nos Estados Unidos, de acordo com Berber Sardinha (2004a, p. 299) é Douglas Biber, pois “impulsiona um tipo de metodologia de estudo de linguagem baseada em corpus chamado Análise Multidimensional”, que foi utilizado como modelo metodológico para o nosso trabalho.

Na esfera metodológica portanto, nossa pesquisa adotou a análise multidimensional (BIBER, 1988; PACHECO DE OLIVEIRA, 1997; SHERGUE, 2003; BERBER SARDINHA, 2000 a, 2004a, 2004b; KAUFFMANN, 2005). Essa abordagem baseia-se no pressuposto de que a coocorrência de um conjunto de características

---

<sup>6</sup> De agora em diante LC

linguísticas em diferentes textos de maneira persistente evidencia um padrão de uso de cunho funcional, ou seja, uma dimensão de análise.

Segundo Biber (1988, p.14), do ponto de vista linguístico “cada dimensão compreende um grupo independente de características linguísticas coocorrentes, e cada padrão coocorrente pode ser interpretado em termos funcionais.” Essas dimensões serão empiricamente avaliadas em termos quantitativos (quantas existem) e relacionais (como se relacionam umas com as outras). Assim, a metodologia quantitativa apoia-se na contagem de frequência de características lexicais e situacionais para estabelecer padrões de ocorrências que possibilitam generalizações a respeito da língua.

O autor defende a ideia de que a análise textual, que considera apenas uma característica lingüística, não dá conta das diferentes situações comunicativas presentes num texto, fazendo-se necessária uma investigação, relacionando as diferentes dimensões (BIBER, 1988, p.12-13).

Dessa forma, num conjunto de textos que se pretende analisar multidimensionalmente (mais de 6.000 textos em nosso caso), são determinadas as características linguísticas de cada texto, computadas suas ocorrências individuais, recorrências por texto e relacionas texto a texto. Essa análise tem cunho estatístico e necessita do auxílio de programas computacionais específicos que dão conta de grande quantidade de dados. Esse contraste evidenciará as dimensões de análise, conforme veremos mais detalhadamente nos capítulos teórico e metodológico.

Outra área que oferece suporte para nossa pesquisa é a de estudos musicológicos associados e sociólogos. Como dissemos no início desta seção, a categorização dos gêneros musicais tem sido proposta por pessoas envolvidas com a mídia e a indústria fonográfica (gravadoras, radialistas, críticos de música em revistas especializadas), pois, como nos informa Moore (2003, p. 1-2), os estudiosos de musicologia, até o final do século vinte, consideravam a música popular como amadora (“apesar do evidente profissionalismo exibido por seus praticantes”<sup>7</sup>). Apenas agora, no início do século XXI, o mundo acadêmico (entendido aqui pelos cursos de música) começou a estudar música popular e seus diferentes aspectos, sociais, o processo de criação e produção, o mercado, entre outros.

---

<sup>7</sup> Do original: “notwithstanding the evident professionalism exhibited by its practitioners”

Os sociólogos, quando estudam música, em geral procuram traçar associações entre música e identidade e pós-modernidade (FRITH, 1993; GRIFFITHS, 2003; MOORE, 2003; STRAW, 2003; HALL, 2006). Na área da musicologia as letras de música recebem pouca atenção, sendo normalmente valorizados os aspectos musicais, composicionais e de produção musical (WISNIK, 1989; MIDDLETON, 1999; MUNDY, 1999; BRACKETT, 2000; STARR & WATERMAN, 2007). Na Linguística as pesquisas geralmente privilegiam investigações sobre características poéticas, semióticas ou exclusivamente temáticas das letras das músicas (NESTROVSKI, 2000; WINKLER, 2000; TATIT, 2001; COSTA, 2002; ROSSATO, 2002; DAMAZO, 2004; CROSSLEY, 2005; MATTE, 2005); ou utilizam-se de letras de música como objeto de ensino (MURPHEY, 1990; MORA, 2000; ARLEO, 2000; BÉRTOLI-DUTRA, 2001; SCHOEPP, 2001).

Acreditamos que ao caracterizar letras de música popular como tendo componentes da linguagem falada coloquial, a música deixará de ser vista como um tipo de ‘poesia cantada’ apenas, ou como um tipo banal de linguagem comercial, da mídia de massa globalizada, passando a ser vista como fonte de língua autêntica. Assim como sugere Griffiths (2003, p. 42-43) “o que precisa acontecer, parece-me, é, primeira e crucialmente, que paremos de pensar que as palavras nas músicas pop são poemas, e comecemos a dizer que elas são *como* poesia, de algumas maneiras, e, por conseguinte, se elas não são como poesias então elas tendem a ser *como* prosa (...) Eu acho que nós podemos começar a pensar sistematicamente sobre o trabalho das palavras e historicamente sobre como seu posicionamento nas canções pop criou, o que pode ser chamado de ‘consciência-sobre-a-palavra’”<sup>8</sup>.

A seguir, apresentamos como este trabalho encontra-se dividido.

## 5. Organização da tese

O presente trabalho apresenta-se distribuído da seguinte maneira:

---

<sup>8</sup> Do original: “What needs to happen, it seems to me, is, first and crucially, that we stop thinking that the words in pop songs are poems, and begin to say that they are *like* poetry, in some ways, and that by extension if they are not like poetry then they tend towards being *like* prose (...) I think that we can then begin to think systematically about the words work, and historically about how their position in the pop song has developed, what might be called ‘word-consciousness’”

Em primeiro lugar, apresentamos um capítulo sobre música e seu papel na cultura, na sociedade e na língua, acompanhado de uma descrição histórica e estilística da música popular anglo-americana.

O capítulo seguinte discute mais profundamente as questões que fornecem aporte teórico para esta pesquisa, ou seja, a Linguística de Corpus e os aspectos teóricos da metodologia de Análise Multidimensional.

Em seguida, apresentamos o estudo-piloto que contribuiu como suporte para o desenvolvimento da pesquisa como um todo, no qual procuramos traçar convergências entre a linguagem da música popular anglo-americana e a língua inglesa em geral.

No capítulo quatro apresentamos os procedimentos metodológicos de nossa pesquisa principal, os quais envolvem a descrição dos corpora usados para a pesquisa, os passos da análise multidimensional: a seleção das variáveis linguísticas observadas e a extração de fatores estatísticos.

No quinto capítulo apresentamos a interpretação dos fatores que nos levaram à definição das dimensões de análise, além disso, discutimos os resultados de nossa análise, apresentando as dimensões de variação na música popular anglo-americana por nós detectadas.

Finalmente, no último capítulo apresentamos a conclusão sobre a pesquisa e nossas considerações finais.

## **1. Música e identidade cultural, sociedade e língua**

Esta pesquisa esteve enraizada na música desde sua concepção, a fim de definirmos o que poderia integrar ou não nosso corpus de estudo, passamos a considerar a música sob diversos aspectos: o gosto popular, o gosto pessoal, diferentes características melódicas e linguísticas, e, especialmente, procuramos desvendar as características que fazem a música tão presente em nossas vidas. Mais especificamente, nossa pesquisa focaliza a música composta e gravada originalmente em língua inglesa, cujas gravações levaram seus intérpretes ao sucesso.

Dessa forma, primeiramente apresentamos a relação da música com identidade cultural. Em seguida, discorreremos sobre a perspectiva de música popular adotada em nossa pesquisa, apresentando breves comentários sobre trabalhos de pesquisa desenvolvidos na área. Continuamos o capítulo com uma perspectiva histórico-estilística da música popular americana, atrelada à influência recebida por outros países, vez que adotamos o ponto de vista de divulgação em massa cuja fonte de partida é do mercado americano para o resto do mundo.

### **1.1. Música e identidade**

A música está atrelada, sociologicamente, à identidade cultural de um povo. Segundo Frith (1993), ao observamos o modo de produção da música veremos que o homem constrói uma experiência – musical-estética – que reflete a sua identidade. Seu principal argumento é que identidade é um processo móvel e a compreensão da experiência musical das pessoas, o fazer e ouvir música, estão relacionados à ética, estética, à performance histórica, individual e social do indivíduo. Para ele, identidade é um processo experiencial que pode ser compreendido pela música, que oferece sentido para o relacionamento do “eu” e do “outro”, “do subjetivo e do coletivo”. Para Frith, música não é uma maneira de expressar idéias, mas de vivê-las.

Outro argumento interessante de Frith (1993) é que a música, enquanto narrativa, texto, concede dinamismo e performance às artes, pois enfatiza o emocional engajando a audiência com a banda e, por isso, não deve ser tida como representante de valores, mas como incorporadora. Ouvir uma canção leva o indivíduo a uma resposta emocional, a estabelecer “alianças” com o *performer*, de forma individual (nossos corpos absorvem/sentem o ritmo) e coletiva (de acordo com gosto, contexto social). A identidade refletida pela música é aquela que gostaríamos de ter, muito mais do que aquela que temos, pois nos identificamos com nossos desejos. Enfim, segundo o autor, a música nos fornece um experiencial daquilo que o ideal poderia ser. Constrói nosso senso de identidade através de experiências que nos inserem em narrativas culturais imaginárias (FRITH, 1993).

O mesmo foi observado por Costa (2001), ao defender a tese de que o discurso lítero-musical é constituinte de identidades, mobilizando o imaginário da sociedade. Descreveu estilos e fases da música popular brasileira, considerando o plano musical, o verbal e o investimento ético e enunciativo dos autores, compositores e intérpretes na formação dessa identidade. Ao longo desse processo, o autor destacou como a música estrangeira entrou em nosso país, na maioria das vezes, desde o início, preservando o estilo, mas cantada em português (id., p. 306), especialmente colaborando para a origem do rock nacional.

Se num determinado momento a música internacional (especialmente a cantada em língua inglesa) colaborou para a formação musical brasileira em língua portuguesa, em outro, invadiu (e continua a invadir) as estações de rádios, as trilhas sonoras das novelas etc., contribuindo para a formação de uma nova identidade mais globalizada – ainda que orientada pelos interesses da indústria musical. Vale ressaltar que, mesmo cantadas em português, essas primeiras canções, originais ou versões, já apresentavam palavras estrangeiras como *baby*<sup>9</sup> e *yeah*<sup>10</sup>, apontando para uma fusão sonora e linguística que se

---

<sup>9</sup> “... Baby, baby/não adianta chamar, oh não/ quando alguém está perdido/ procurando se encontrar/ Baby, baby/ não vale a pena esperar, oh não/ tire isso da cabeça/ ponha o resto no lugar” (canção “Ovelha Negra”, composta e gravada por Rita Lee, em 1975)

<sup>10</sup> “*splish splash!*/ fez o beijo que eu dei/ nela dentro do cinema/todo mundo olhou-me condenando/só porque eu estava amando/ agora lá em casa/ todo mundo vai saber/ que o beijo que eu dei nela/ fez barulho sem querer/ yeah!” (canção “*Splish Splash*”, gravada por Roberto Carlos, em 1961, versão em português da original de Bobby Darin/Jean Murray)

mantém até hoje.

Como podemos ver, no mundo pós-moderno, ou da modernidade tardia (cf. HALL, 2006), globalizado, as diversas identidades aparentemente sofrem ameaça de fragmentarem o indivíduo moderno com tantas “invasões”. Na realidade, a questão do conceito de identidade cultural é bastante questionável. Segundo Hall (2006, p.8), a identidade está sempre em formação ao longo do tempo, por meio de experiências reais imaginárias e deveria ser vista como um ‘processo em andamento’ chamado identificação. Historicamente, Hall lembra a evolução do sujeito iluminista (individualista) para o sociológico (influenciado por mundos culturais externos) e para o sujeito pós-moderno (fragmentado, móvel, sem uma identidade fixa permanente). Este último, caracterizado essencialmente pelas rápidas e constantes mudanças do mundo. Ressalta o valor da cultura nacional (local) na formação de uma identidade nacional do sujeito, pois produz valores com os quais nos identificamos representados pelos “vínculos a lugares, eventos, símbolos, histórias particulares” (Ibid., p. 76).

Ainda segundo Hall (2006, p. 14), “as sociedades modernas são sociedades de mudança constante, rápida e permanente. Esta é a principal distinção entre as sociedades ‘tradicionais’ e as ‘modernas’”<sup>11</sup>. Ainda assim, algumas pessoas sentem a sua cultura ameaçada pelo processo de globalização e não o veem como uma evolução histórica e culturalmente natural, isto é, sentem que a sua condição de pertencimento está em risco, ignorando o hibridismo cultural e étnico de quase todas as nações modernas, ao fazer uma nova combinação entre tempo e espaço, interconectando comunidades e atravessando fronteiras. Porém, as trocas culturais entre as nações e o consumismo global devem estar criando possibilidades de “identidades partilhadas”, o que não significa enfraquecimento cultural, mas a possibilidade de se fazerem escolhas quanto aos elementos (culturais) com os quais cada um mais se identifica. Isso permite que o indivíduo de um país escolha, entre todas as opções disponíveis no mundo e, dessa forma, compartilhe a identidade que escolheu com os que compartilham seu mesmo gosto, em qualquer parte do mundo, sem necessariamente ter que sair de casa. Enfim, a globalização deve produzir novas identidades, locais e globais, ainda que a

---

<sup>11</sup> Aspas do original

troca cultural<sup>12</sup> aconteça de forma desigual<sup>13</sup>, causando no sujeito a possibilidade de “contestar e deslocar as identidades centradas e ‘fechadas’ de uma cultura nacional”(HALL, 2006, p. 87).

No caso da música, a identidade escolhida por um sujeito em um determinado país pode ser compartilhada por outro sujeito do outro lado do mundo. O que faz a música especial para a identidade é que ela é a forma cultural que melhor capacita-nos a “cruzarmos as fronteiras”, pois o som nos carrega através das cercas, muros e oceanos, através das classes, raças e nações; define lugares (clubes, cenas, e *raves*, com fones de ouvido, rádio ou em salões de concertos). “Nós apenas vamos onde a música nos leva” (FRITH, 1993, p. 126). A título de ilustração, podemos observar em qualquer parte do mundo membros da nação cabeluda<sup>14</sup>, apreciadora do estilo metal pesado, que escolheram essa identidade musical que se transferiu para a moda e comportamento social (roupa, cabelo, linguagem, etc.) que reconhecem a si mesmos e são facilmente identificáveis; assim como os ‘*country*’ (estilo *cowboy*, chapéu, jeans, bota, cinturão); os góticos (roupas e cabelos extremamente pretos); os ‘*emos*’ (cabelos curtos, com franja longa cobrindo o rosto, roupas escuras, ar de tristeza); os psicodélicos (roupas soltas e coloridas, pés no chão, barba), entre outros.

Procuramos mostrar até agora que, embora detentores de uma identidade nacional conjunta, há diferentes identidades individuais que podem estar ligadas a fatores estrangeiros à nossa cultura. A música é um desses fatores, não só por sua característica sonora mas também pela fusão estética, comportamental, linguística e textual (FRITH, 1993). Tais características não podem ter sido transferidas exclusivamente pela música. Não haveria como uma pessoa resolver deixar os cabelos mais longos simplesmente por ouvir aos Beatles nos anos sessenta. A imagem

---

<sup>12</sup> Obviamente devemos considerar que o mercado de consumo e as imposições da mídia direcionam o consumidor, ou seja, existe o fato irrefutável de que a penetração do estrangeiro acompanha as mudanças mercadológicas e, conseqüentemente, linguísticas, como é o caso do inglês ser Língua franca hoje. (conf. Frith, 1998)

<sup>13</sup> O que na verdade sempre ocorreu, pois nossa sociedade ocidental foi formada por guerras, repatriações, massificação de telecomunicação, etc. O autor até cita que a importância da calça jeans para quem mora em Taiwan ou China como sendo a mesma para o ocidente.

<sup>14</sup> Hair nation, hair-metal, hair bands, todas expressões que simbolizam os fanáticos pelo estilo de música metal dos anos 80. “... is a term used to describe the visual style of certain heavy metal music bands that arose in the late 1970s and early 1980s in the United States... ‘Hair bands’ was the term popularized by MTV in the 1990s and derives from the tendency among most such bands to style their long hair in a teased-up fashion.” ([http://en.wikipedia.org/wiki/Glam\\_metal#cite\\_note-0](http://en.wikipedia.org/wiki/Glam_metal#cite_note-0))



transferida por meio de fotos, da televisão e dos filmes colaborou para isso, entre outros fatores sociais, políticos, mercadológicos, etc. Ou seja, música e mídia estão muito ligadas à formação da identidade de um sujeito. Por isso, passaremos a discutir brevemente a influência do vídeo associado à música.

## 1.2. Música e vídeo

Will Straw (2003) coloca a música no cenário pós-moderno levantando semelhanças, dessemelhanças e implicações históricas e sociais sobre o relacionamento da música rock com o vídeo musical. Segundo o autor, depois do advento da MTV, a música não pode mais ser considerada apenas por suas características sonoras e linguísticas, mas também pelo conteúdo semiótico.

O autor aponta o surgimento da MTV, em 1981, como marco transformador da música, não apenas para os EUA, mas para o mundo. Antes da MTV (que era uma espécie de rádio nacional visual nos Estados Unidos), a venda da música, mais especificamente a rock branca, levava muito tempo, pois baseava-se nos lançamentos dos álbuns dessas bandas, entre os quais havia um grande hiato temporal, o que fazia com que o público alvo (jovens adolescentes, que compram discos) envelhecesse e acabasse, em parte, sendo perdido.

A MTV concedeu maior velocidade comercial e cultural ao mundo da música. Unindo-se ao estilo dance da música negra, estabeleceu a comercialização dos singles (agora em formato vídeo-clipe), que frequentemente serviam de testagem antes do lançamento oficial de um disco ou banda e, secundariamente, realçou a relação da identidade do performer com a mitificação da celebridade<sup>15</sup> – o que contribuía para a vendagem, mas pouco se relacionava com a “interpretação ou a compreensão da música em si” (STRAW, 2003, p. 11). Porém, essa imagem integrava a “riqueza semiótica e o alto nível de contextualização” (P. 10) que dotou a música popular desse período, fornecendo maior autonomia de discurso e possibilitando aos ouvintes liberdade para selecionarem suas identificações.

---

<sup>15</sup> contempladas pelos programas e revistas que enfatizam sua biografia

Goodwin (2003) ressalta que a MTV pode ser considerada como texto pós-moderno por ser a emissora que carregava dentro de um padrão formal de televisão (ou rádio) questões ideológicas frequentemente relacionadas ao mudo do *rock*, do *art-rock* e do *pop* (inicialmente, branco<sup>16</sup>), do *heavy metal* e do *rap*, incentivando a busca por direitos, apoiando causas políticas, ambientais e raciais<sup>17</sup>. O envolvimento da música, especialmente do *rock*, com esse tipo de manifestação, segundo Szatmary (2000), foi possível, nos EUA, a partir do momento que a música branca se encontrou com a negra. Os negros foram os responsáveis por incrementar os sons com guitarra e ritmos e o *rock* ajudou a integrar a América branca à negra. Segundo o autor, essa música refletiu os ideais e pensamentos das gerações pós-guerra, a influência do *punk* britânico na música e na vida daquele povo, a ideologia da geração baby-boomers<sup>18</sup>, ressaltou a necessidade de se considerar os problemas do povo do campo, da guerra do Vietnam, e continuou a refletir a vida do povo, até mesmo a revolta da geração dos anos noventa com suas vidas advindas de lares desfeitos, os quais foram nitidamente refletidos na violência do som *hardcore* e *grunge*.

Segundo Berland (2003), o vídeo musical integrou fortemente música/filme/televisão garantindo seu poder global e local de entrar no espaço das pessoas e de praticamente impossibilitar que uma música faça sucesso sem um vídeo. Embora o vídeo musical não exista sem a música, paradoxalmente ele nos atrai e nos repele da música por seu apelo visual. Para a autora, as imagens da cultura pop, não apenas dos vídeos, representam um movimento histórico que integra o imaginário dos sujeitos, assim como proposto por Frith (1993). Entre essas imagens está a do rock star que deixa de ser um mero músico quando entra para o cinema (1957)<sup>19</sup>, contribuindo para a dissociação entre som e imagem e, mais tarde, para sua reunificação, por meio das exigências mercado-tecnológicas, especialmente, da televisão.

De acordo com Berland (2003, p. 36) enquanto a música preenche o espaço que circunda o sujeito, funcionando como uma espécie de “extensão do corpo em direção ao

---

<sup>16</sup> Interessante que todos os trabalhos por nós consultados comentam as críticas de uma MTV exclusivamente branca, no início, que mudou com o sucesso, especialmente, de Michael Jackson.

<sup>17</sup> *Black Rock Coalition, AIDS, Rock Against Drugs*, entre outros.

<sup>18</sup> *Baby Boomers* representa a geração que nasceu após o final da Segunda Guerra Mundial. Nessa época houve, nos EUA, um grande aumento na taxa de natalidade e, em 1959, a população adolescente representava 30% da população total americana, atraindo a atenção para o público jovem enquanto consumidores. (GILLON, 2004).

<sup>19</sup> A entrada de Elvis para o cinema.

social, e vice-versa, a televisão tenta envolver” o sujeito “para atrair seus olhos para um ponto único e fixar o resto dele diante dela”. No vídeo, o cantor, ‘poeta lírico’ que responde à narrativa visual com sensatez, é o mediador entre câmera (imaginário) e audiência (real). O ritmo emoldura o encadeamento de imagens, mantendo a estrutura verso/refrão/instrumental com naturalidade, antecipando o olhar para depois conquistá-lo. “Nos versos, o cantor descreve, no refrão expressa-se, no intervalo instrumental ilumina sua causa”(p. 40) e sua fantasia, abre-se para a amplitude do mundo. Enfim, a relação som/imagem-televisiva cria um ciclo de reprodução de símbolos, de competência e prazer dependente que se identifica com a juventude.

### 1.3. Música e armazenagem

Outro marco que mudou a face da indústria musical foi, segundo Szatmary (2000), o advento dos CDs. Com eles, artistas antigos retornaram às paradas de sucesso, à vida, pela necessidade dos fans em readquirirem os antigos LPs na forma de CD.

A velocidade de lançamento e armazenamento das músicas aumentou. Para esse autor, isso é indicação de que a música tanto influencia como é influenciada pelo seu público, que ficou mais exigente quanto à distância temporal entre o lançamento de um álbum e outro.

Hoje em dia, com o desenvolvimento de arquivos comprimidos e formato MP3 e MP4, as músicas são facilmente armazenadas e transferíveis. Ou seja, se na época do disco de vinil os amantes de música passavam horas gravando fitas com faixas musicais advindas de diferentes discos (LP) e na época dos CDs isso também era feito da mesma forma, mas digitalmente. Hoje, a fita que gravava 60 ou 90 minutos e o CD que gravava até 4 horas de música foram substituídos por arquivos digitais, gravados em “*pen drives*” ou aparelhos que reproduzem arquivos em MP3<sup>20</sup> que tocam horas e horas a fio, podem ser carregados no bolso da camisa e não exigem que ninguém tenha (necessariamente) pago pela música, pois há programas computacionais que oferecem compartilhamento de arquivos, ou seja, qualquer um pode baixar qualquer música da internet. Starr e Waterman (2007, p. 470), comentam que a “música popular tem sido

---

<sup>20</sup>Arquivo digital que comprime arquivos sonoros, uma variação da extensão MPEG – usada originalmente na tecnologia para DVD.

catalisada pela internete”<sup>21</sup>, e o surgimento de aparelhos como o IPOD<sup>22</sup> tem possibilitado que os “consumidores de música popular estejam conectados a uma matriz de entretenimento global que inclui computadores pessoais, a Internet, serviços de descarregar [baixar, fazer *download*] músicas como o novo Napster e os serviços que estão começando a substituir as funções tradicionais de divulgação”<sup>23</sup> (p. 472).

A fim de exemplificar o desenvolvimento tecnológico, os autores também citam a crescente prática de “*podcasting*”, ou seja, a distribuição digital de arquivos de imagem e vídeo, que podem até mesmo ser visualizados em alguns telefones celulares. Isso sem falar em sítios da Internete como o *youtube.com*, onde se pode ver desde vídeos pessoais, caseiros, até antigas (e as mais recentes) gravações vídeos-musicais.

#### 1.4. Música e linguagem

Como vimos, a música popular está fortemente ligada a questões sociais, culturais, econômicas e tecnológicas, além do gosto (identidade). Frith (1998) ressalta que o gosto não pode ser medido, nem classificado, mas moldado pela música. Para Ana Cristina Matte (2005), que explorou o gosto pela música em sua pesquisa, a identificação do ouvinte com determinado estilo musical não é o único fator de influência, mas defende a consideração especial de questões de poder (classe social), saber (acesso à cultura), querer (faixa etária) e dever (religião). Segundo a autora, a música expressa emoções e materializa os elementos da paixão culturalmente determinados. Tanto para Matte quanto para Frith, a música envolve sentimento e independe de classificação como boa ou ruim (o que vai depender dos padrões considerados por quem classifica: classe social, valores culturais, valores religiosos), pois seu valor reside na semiose de seu significado com a forma/modo como é dito, e deve variar de acordo com o público alvo.

Interessa-nos, em especial, nesta questão de identidade cultural, o papel da língua na música. Se considerarmos que música pode refletir essa condição de pertencimento

---

<sup>21</sup> Do original: “popular music has been catalyzed by the internet”.

<sup>22</sup> Reprodutor portátil de som e imagem comprimidas.

<sup>23</sup> Do original: “contemporary consumers of popular music are connected to a global entertainment matrix that includes home computers, the Internet, music download services such as the new Napster, and the new services that are beginning to supplant the traditional functions of broadcasting”.

de um povo, estaremos conseqüentemente considerando suas raízes na cultura e na história daquele povo, transmitidas não somente pelo som, como também pela letra e pela imagem da música. Assim, podemos inferir que conhecer a música de um povo é também conhecer sua cultura, sua história e sua língua. Hall (2006) vê a língua da seguinte forma:

*A língua é um sistema social e não um sistema individual. Ela preexiste a nós. Não podemos, em qualquer sentido simples, ser seus autores. Falar uma língua não significa apenas expressar nossos pensamentos mais interiores e originais; significa também ativar a imensa gama de significados que já estão embutidos em nossa língua e em nossos sistemas culturais. Além disso, os significados das palavras não são fixos, numa relação um-a-um com os objetos ou eventos no mundo existente fora da língua*

(HALL, 2006, p. 40)

A música, na concepção de Matte (2005) é uma linguagem semissimbólica que se diferencia da linguagem verbal por não se constituir pelos signos, mas caracterizar-se semioticamente no plano da expressão e do conteúdo. Considera a música como objeto partilhado, não acumulável e inexistente fora do tempo.

Já o componente semiótico transferido pelos gestos e pela voz do cantor, segundo Luiz Tatit (1996; 2001), proporciona certa empatia entre artista e público, um se colocando um pouco no lugar do outro. Já o componente semiótico presente no texto – na letra da música – se manifesta no conteúdo, ou seja, na superfície do texto reside a compreensão de valores semióticos, como o espacial, o temporal, os de ordem tensivas e conceituais e valores de significação que estão expressos num “sistema complexo de funções sintáticas” (2001, p.15)

Middleton (2000, p.28-29) lembra-nos que a música não simplesmente “funciona como (isto é, por imitação de) discursos verbais, mas que tanto música quanto linguagem habitam um conjunto maior”.<sup>24</sup> Em outras palavras, o autor quer dizer que as emissões discursivas presentes na música e fora dela encontram um ponto de intersecção, que envolve, entre outras características, acordos tácitos entre artistas e fãs quanto à compreensão da subjetividade e estruturas presentes na música e seus efeitos.

Muitos estudos consideram os aspectos poéticos de letras de música, pois é incontestável a relação entre poesia e música, vez que partilham “recursos semelhantes ao processo de criação poética, quais sejam, a métrica, o sentido figurado, a rima”

<sup>24</sup> Do original: “meant not that music functions *like* (that is, in imitation of ) verbal discourses, but that both music and language inhabit a larger set”

(COSTA, 2002, p.112). Por outro lado, seu componente melódico distancia as letras de música do *status* ocupado pela poesia no discurso literário. O autor considera poesia e letras de música como gêneros diferentes “que se interseccionam por aspectos de sua materialidade e por alguns momentos comuns de sua produção” (id., *ibid.*, p.113). A abordagem poética da letra de música não é o foco deste trabalho, pois não estamos investigando temas, rimas ou conteúdos metafóricos, mas sim os componentes léxico-gramaticais das letras.

As letras de música, segundo Frith (1998) integram o gênero coloquialismo conversacional, apresentando inúmeras variações linguísticas, não devendo ser consideradas simplesmente como poesias cantadas, mas uma forma de retórica e oratória. “As melhores canções populares, resumindo, são as que podem ser ouvidas como uma *luta* entre a retórica musical e a verbal, entre o cantor e a canção”<sup>25</sup> (FRITH, 1998, p. 182). Essa relação leva os ouvintes a fazer alianças afetivas e emocionais<sup>26</sup> que ajudam a construir a identidade social. Por meio das experiências corporais, temporais e sociais que a música oferece, ela insere a pessoa no contexto imaginário de narrativas culturais.

Além disso, a música não acontece em esferas separadas, ou seja, melodia, letra, performance do cantor, etc., mas, tudo ao mesmo tempo, como coloca Middleton:

*Se a música deve ter seu próprio momento discursivo irredutível (embora nunca puro), as correspondências, equivalências e paralelos que os padrões sonoros sugerem residem frequentemente não na esfera da linguagem, mas na dos gestos, do processo somático e da sensação tátil.*<sup>27</sup>

(MIDDLETON, 2000 p. 11)

Costa (2002, p. 107) adverte que a música – que ele chama de canção por referir-se à letra e melodia – “é um gênero híbrido, de caráter **intersemiótico**, pois é resultado da conjugação de dois tipos de linguagens, a verbal e a musical”<sup>28</sup> e, portanto, exige

<sup>25</sup> Do original: “The best pop songs, in short, are those that can be heard as a *struggle* between verbal and musical rhetoric, between the singer and the song”. Itálico do original.

<sup>26</sup> Tais alianças também existem em outras áreas da cultura popular como esporte e moda

<sup>27</sup> Do original: “If music must be granted its own irreducible (though never pure) discursive moment, the correspondences, equivalences, and parallels that its sound-patterns suggest lie often not in the sphere of language but in that of gesture, somatic process, and tactile sensation”

<sup>28</sup> itálico e negrito cf. original

que apliquemos três competências para sua análise: verbal, musical e lítero-musical (articulação entre verbal e musical).

Em geral, os estudos de musicologia concentram-se nas características exclusivas da composição musical, pois os musicólogos, geralmente possuem pouca afinidade com estudos de palavras, concentrando-se em convencer os colegas do valor da música popular pela explicitação de características musicais e, conferindo pouca atenção aos aspectos textuais. De acordo com Middleton (2000), os primeiros estudos musicológicos sobre música popular basearam-se num modelo de análise descritiva e estrutural, da mesma maneira tradicional que se estudava música clássica, o que provavelmente causou muita má interpretação, provocando estudos que falavam “muito mais sobre os autores do que sobre a prática musical de cultura popular”<sup>29</sup> (id. Ibid., p. 3). Tais estudos, na realidade, estavam centrados nas características musicais, tais como compasso, harmonia, ritmo, qualidade sonora, etc. Estudos mais atuais consideram fatores como consumismo; populismo; a influência do discurso (valor estético, formação política, identidade histórica, gênero, sexualidade, etnia, etc.) na produção de significados; a especificidade dos discursos musicais (gênero, técnica e tecnologia musical; retórica; história do estilo; linhagem e influência, etc.) e assim por diante. Existem ainda, autores que exploram o potencial dialógico da música, à luz de Bakhtin. Segundo Middleton (2000), esses estudiosos entendem que os significados das canções “são produzidos por meio de diálogo em muitos níveis: entre as texturas, vozes, estruturas e alianças estilísticas de cada evento musical; entre produtores e público alvo; entre texto, estilo e gênero e outros textos, estilos e gêneros; entre discursos, musical e outros; entre interpretações, mediadores e outros atores sociais envolvidos”<sup>30</sup> (p.13).

Todavia, encontramos alguns trabalhos que observam as letras também, como é o caso de Peter Winkler, que analisa a letra da música *I think it's going to rain today*, de Randy Newman. Apresentando uma análise bem detalhada, verso por verso, Winkler (2000), conclui que a letra usa vocabulário bem simples que disfarça a ironia presente, ou seja “aquilo que não é dito é frequentemente mais importante do que aquilo que é

<sup>29</sup> Do original “... we were told more about the writers than about the musical practices of popular culture”

<sup>30</sup> Do original “... are produced through dialogue at many levels: with the textures, voices, structures, and style-alliances of the individual musical event; between producers and addressees; between text, style and genre and other texts, styles, genres; between discourses, musical and other; between interpretations, mediators, and other involved social actors.”

dito”<sup>31</sup> (p.28). A música, segundo o autor, inicia oferecendo imagens de abandono, ausência e descontinuidade<sup>32</sup>, que continuam ao longo da letra, sendo reforçadas por momentos de tensão pela situação retratada pela letra – a decadência das cidades americanas, da família, o descaso para com as pessoas necessitadas, que passam fome e desespero (p.56). Winkler (2000, p.31) traça também uma relação entre a letra e a maneira que o cantor canta, na qual a voz de Newman reflete o tom de amargura e tristeza, associada a uma melodia agitada que, ao destoar da letra, aumenta a sensação de angústia exacerbada por um “grito: *‘lonely, lonely’*”. Sua conclusão nos leva a duvidar do sucesso que tal música possa ter conquistado. Interessantemente, essa música, gravada em 1968 originalmente, foi regravada diversas vezes por artistas como Alan Price, Dave Van Ronk, Judy Collins, the Everly Brothers, Claudine Longet, Dusty Springfield, Nina Simone, Pat Boone e Peggy Lee.

Uma análise como esta leva em consideração os diferentes aspectos semióticos da música e a mensagem que ela quer passar. Diversas pesquisas, na área de letras, aparecem considerando os aspectos poéticos das letras de música, como, por exemplo as pesquisas de Rossato (2002) sobre Caetano Veloso e Legião Urbana; e de Damazo (2004) que explora os atributos poéticos das canções de João do Vale. Outras realizam estudos sobre características linguísticas específicas presentes na música. Crossley (2005), por exemplo, analisou as metáforas detectadas na música afro-americana, defendendo que essas metáforas são usadas para expressarem as experiências, ideias e significados mais complexos de uma forma mais acessível para a compreensão do público.

Existem também, diversos estudos que exploram o uso da música para o ensino de línguas e de línguas estrangeiras, como o de Schoepp (2001), por exemplo, que enfatiza as características afetivas e cognitivas da música para o ensino de línguas; ou como os de Murphey (1990) e Mora (2000) que ressaltam as características afetivas e sonoras da música para o ensino. Poucos, contudo, enfocam a música como fonte central para a aprendizagem de língua estrangeira (L2), como fizemos em nossa pesquisa de mestrado (BÉRTOLI-DUTRA, 2002) ao desenvolvermos atividades de aprendizagem de L2 por meio do reconhecimento e reprodução de padrões linguísticos

---

<sup>31</sup> Do original:” what is left unsaid is often more important than what is said”

<sup>32</sup> “broken windows and empty hallways/ pale dead moon in a sky streaked with grey”



encontrados em letras de música. Entendemos que trazer a música para sala de aula, se por um lado pode ser fator de grande motivação, ao mesmo tempo cria bastantes dificuldades, pois esta questão envolve o preparo constante de materiais, por parte do professor, uma vez que as músicas quando integrantes do livro didático podem ficar datadas e/ou sair do nicho de interesse dos aprendizes.

As pesquisas mencionadas até aqui servem para confirmar que esta área é bastante explorada, contudo, vale lembrar, mais uma vez, que o enfoque de nosso trabalho não está nos aspectos poéticos, temáticos ou pedagógicos das letras de música, conseqüentemente, características como as observadas pelos pesquisadores citados não fazem parte do escopo de nossa análise.

Passaremos, a seguir a discutir música popular seguida de apresentação sobre a trajetória da música anglo-americana que é foco de nossa pesquisa. Faz-se necessário ressaltar que, como nosso trabalho tem foco em música popular de língua inglesa, a maioria dos estudos observados foi desenvolvida por estudiosos cuja língua materna também é a inglesa. O Brasil possui vasta literatura sobre estudos de música popular que geralmente focalizam aspectos particulares da música popular brasileira, como os movimentos da *tropicália* e *bossa nova*, por exemplo, que não são objetos desta pesquisa e, portanto, não foram considerados.

## 1.5. Música Popular

É difícil chegar a uma definição que satisfaça a todos os apreciadores de música. A noção de que cultura de massa não tem valor acadêmico, estético ou intelectual, contribuiu para a disseminação de um preconceito contra as diversas formas de arte popular. Bennett (2007) aponta este problema citando a posição do teórico em cultura Dwight MacDonald:

*Cultura de massa vem imposta de cima. É fabricada por técnicos contratados por empresários; sua audiência é de passivos consumidores, sua participação é limitada à escolha entre comprar ou não comprar. Os senhores do kitsch<sup>33</sup>, em*

---

<sup>33</sup> Nota nossa: de acordo com o dicionário Aurélio, Kitsch significa objeto ou estilo que, simulando obra de arte, é apenas imitação de mau gosto para desfrute de um público que alimenta a indústria da cultura de consumo ou cultura de massa; atitude ou reação desse público em face de obras ou objetos com essa característica.

*resumo, exploram as necessidades culturais das massas a fim de obterem lucro e/ou manterem a ordem de sua classe (1957[1953]:60).*<sup>34</sup>

(MacDONALD, 1957, p. 60 apud BENNETT, 2007, p. 8)

O termo popular é definido pelo dicionário Aurélio (FERREIRA, 1999) da seguinte forma: “adj. Que pertence ao povo; que concerne ao povo. / vulgar; plebeu / que desperta a simpatia, o afeto do povo. / muito conhecido, notório.” As duas primeiras definições parecem estar ligadas ao sentido depreciativo da palavra, ou seja, aquilo que é do povo, é vulgar, sem valor. As outras duas parecem carregar um sentido mais positivo que ao ser atrelado à música, pode se referir a quantidade e vendagem de um álbum, CD, etc. Sob essa perspectiva, Middleton (1990) nos apresenta duas abordagens interpretativas do termo música popular.

A primeira, chamada positivista, focaliza o sentido quantitativo da palavra popular, ou seja, “muito conhecido”, considerando música popular aqueles “itens musicais mais amplamente disseminados pela mídia de massa<sup>35</sup>” (MIDDLETON, 1990, p. 5). Essa abordagem apresenta problemas no que diz respeito à quantidade, uma vez que está sujeita a “manipulação e corrupção” (i.e. a difusão está sujeita, por exemplo, aos chamados “jabá” – quando a gravadora paga para a música aparecer mais vezes). Outros problemas estão relacionados a (a) tratar mercados heterogêneos como únicos; (b) tendência de considerar apenas o público jovem (que deixou de ser o único público alvo das gravadoras); (c) foco em gravações não originais, ou seja, ao vivo; (d) ignoram práticas musicais não comerciais; e (e) desconsideram o período de vendagem (quanto tempo levou para que determinada música vendesse tantas cópias).

A segunda abordagem descreve o termo ‘música popular’ considerando seus aspectos sociológicos e existencialistas. Para Middleton (1990, p. 5) “aqui a essência do popular é constante (...) independente se ‘o povo’ é considerado como sujeito histórico

---

34 Do original: “Mass Culture is imposed from above. It is fabricated by technicians hired by businessmen; its audiences are passive consumers, their participation limited to the choice between buying and not buying. The Lords of kitsch, in short, exploit the cultural needs of the masses in order to make a profit and/or to maintain their class rule (1957 [1953]: 60).”

35 Do original “the most widely disseminated items of music disseminated in the mass media”

progressivo e ativo ou vítima manipulada”<sup>36</sup>. Aqui, música popular é aquela de raiz, mas o sujeito é um “receptor vazio”<sup>37</sup>.

Aparentemente, o problema com essas abordagens é que elas estão sempre comparando a música com outra coisa e categorizando-a entre “melhor ou pior; de elite ou de massa; alta ou baixa; aristocrática ou plebeia e assim por diante”<sup>38</sup> (id., p.6)

Posicionando-se contrariamente a essa perspectiva contrastiva, Starr e Waterman (2007) oferecem uma definição ampla de música popular:

*Neste livro usamos o termo ‘música popular’ de maneira ampla, para indicar a música que é produzida em massa e divulgada via mídia de massa; que é diversas vezes ouvida por um grande número de americanos; e que tipicamente conta com uma variedade de tradições musicais pré-existentes. Entendemos que música popular deva ser vista em relação a uma paisagem musical mais ampla, na qual vários estilos, audiências e instituições interagem de maneiras complexas. Este mapa musical não é estático – está sempre em movimento, sempre evoluindo*<sup>39</sup>

(STARR; WATERMAN, 2007, p. 2)

Essa definição é também a adotada nessa pesquisa, pois concordamos com a ideia de que as diversas divisões oferecidas pelos musicólogos colaboraram para a disseminação do preconceito contra o popular. Não nos interessa aqui defender ou atacar, por exemplo quem é ou não popular. Matte (2005), por exemplo, na tentativa de determinar “a boa música”, analisou fala de três artistas e apontou que Chico Buarque sabe compor embora não conheça formalmente música teórica e tecnicamente, sobrepondo o compositor ao instrumentista. Waldick Soriano é bom porque vende muito, valorizando a eficiência à criatividade. Edu Lobo vende mais quando tem um corpo e uma voz que carregam as emoções de suas composições, como Elis Regina, valorizando o intérprete. Conclui, portanto, que não existe a “boa música”, pois seu valor reside na semiose de seu significado com a forma/modo como é dito.

<sup>36</sup> Do original “Here the ‘essence’ of the popular is constant (...), whether ‘the people’ is regarded as an active, progressive historical subject or a manipulated dupe”.

<sup>37</sup> Do original “blank receptors”.

<sup>38</sup> Do original: “better and worse; elite and mass; higher and lower; aristocratic and plebeian, and so on.”

<sup>39</sup> Do original “In this book we use the term ‘popular music’ broadly, to indicate music that is mass-reproduced and disseminated via the mass media; that has at various times been listen too by large numbers of Americans; and that typically draws upon a variety of preexisting musical traditions. It is our view that popular music must be seen in relation to a broader musical landscape, in which various styles, audiences, and institutions interact in complex ways. This musical map is not static- it is always in motion, always evolving.”

Para Starr e Waterman (2007, p. 3-4) a música popular americana é fruto da mistura de diversas características que não podem deixar de ser observadas: o ato de ouvir; a relação entre música e identidade; relação da música com a tecnologia; a relação da música com o comércio musical (o mundo dos negócios); a dicotomia entre a música de centro (*mainstream*) e periferia; e a influência musical advinda de outras culturas.

Em resumo, a música popular segue também a aspectos formais de criação musical (estrutura, boa qualidade sonora, dialeto, letras - se as palavras fornecem acesso imediato ou requerem interpretação); apresenta relação com a identidade dos diversos ouvintes; está atrelada, mas não condicionada, aos avanços tecnológicos dos meios de divulgação; é caracterizada pelo local de produção se central (nos EUA Nova Iorque, Los Angeles e Nashville<sup>40</sup>) ou periférica<sup>41</sup>; faz parte de uma indústria que envolve desde o compositor ao vendedor, ao fã, etc; e é fruto da influência de diversas culturas, mais especificamente a música europeia (baladas inglesas, valsa, galope, polca); a música africana (o sincretismo e as ideias sobre família, igreja, escola, etc., presentes nas letras e as repetições sonoras (*rifes*); texturas polirrítmicas (vários ritmos ao mesmo tempo); variações de timbre; e o *rhythm and blues* (R&B); e a música latina (tango, a *habarena*, a *rumba*, o *mambo*, o samba, a bossa nova, entre outros).

A seção seguinte apresenta um breve panorama da trajetória musical americana relacionada ao que chamam de “invasão” de música advinda de outros países, especialmente a britânica, culminando na divisão dos gêneros midiáticos.

## 1.6. Trajetória da Música Anglo-Americana

Falaremos da história da música popular americana a partir de alguns eventos que marcaram mudanças em sua história, especialmente quanto à divulgação e produção, pois são os meios que tornam a música acessível à massa consumidora. Adotamos como referência principal o trabalho de Starr e Waterman (2007) que atrela a

<sup>40</sup> Ou onde existe poder econômico – “gosto de uma classe média ou alta, urbana, branca e protestante”, segundo Starr e Waterman (2007, p. 9-10).

<sup>41</sup> Gravadoras fora do *mainstream* que produziram artistas afro-americanos; brancos sulistas pobres; classe trabalhadora; imigrantes judeus e latinos; adolescentes; gays (STARR; WATERMAN, 2007, p.9-10).

música ao período temporal e histórico, pois, dessa forma poderemos traçar uma relação entre a época, os artistas e os gêneros musicais de cada período.

Nossos comentários sobre a parte formal da música são breves pois, não possuímos formação musical<sup>42</sup> e o foco de nosso estudo reside sobre os componentes linguísticos. Outro ponto importante é que, embora o corpus de estudo da pesquisa englobe músicas produzidas entre os anos de 1940 e 2009, julgamos pertinente apresentar brevemente o passado histórico da música popular como fonte de embasamento para nossas observações posteriores, especialmente por considerarmos a música popular como resultado de influências que contribuíram, de certo modo, para a caracterização de gêneros que observaremos.

O primeiro marco importante a conferir novos rumos à música popular foi a criação do fonógrafo, em 1877, por Thomas Alva Edison. Antes do aparelho, a música chegava ao público somente por meio de apresentações e por programas de rádio ao vivo; era vendida em formato de partituras que eram reproduzidas em casa, em festas e saraus por aqueles que detinham conhecimento musical. O fonógrafo possibilitou a reprodução e a venda após a morte do cantor/interprete/compositor, criando também um certo estreitamento cultural:

*a música gravada se tornara um meio pelo qual pessoas que nunca se conhecerão face a face comunicam-se através das grandes distâncias geográficas e culturais e , assim, uma importante base para criação (e constante reinvenção) de uma cultura americana distintiva. Qualquer que seja a opinião de alguém sobre esse assunto, não há como negar que a introdução do fonógrafo no final do século dezenove foi um marco de uma época no desenvolvimento da música popular<sup>43</sup>*

(STARR; WATERMAN, 2007, p. 36)

Nessa época, teve início um processo de produção musical conhecido por TIN PAN ALLEY, que produziu os primeiros sucessos de vendagem e as primeiras estrelas da música popular. Os compositores e letristas eram contratados por editoras para

---

<sup>42</sup> Se fôssemos explorar a faceta musical teríamos que escrever outro capítulo somente para explicar os componentes musicais como cadência; suspensão; variações de andamento; ritmo, oitavas, etc. Ainda assim, alguns desses termos serão utilizados e explicados brevemente ao longo do nosso texto.

<sup>43</sup> Do original: “recorded music has become a medium by which people who will never meet face to face communicate with one another over wide geographical and cultural distances, and thus an important basis for creation (and constant reinvention) of a distinctively American culture. Whatever one’s position on these matters, there can be no denying that the introduction of the phonograph during the late nineteenth century was an epochal step in the development of popular music”

“regularmente produzirem novas canções”<sup>44</sup> (STARR; WATERMAN, 2007, p. 29) que eram interpretadas por cantores, chamados de “*crooners*” (somente homens), como Rudy Vallee, Bing Crosby, Russ Colombo (anos 20 e 30) e, mais tarde, Frank Sinatra (desde 1935).

As letras exaltavam a riqueza, a privacidade e o romance, enquanto camuflavam os problemas da época (e.g. Ku klux klan e a grande depressão de 1929). Eram acompanhadas por uma melodia que emprestava os ritmos do jazz<sup>45</sup>, privilegiando o modo verso + refrão AABA, que moldou a música americana da época. O formato AABA consiste numa seção A, que “representa a melodia principal, o padrão básico para a letra”, a tonalidade que a sustenta. Essa seção é repetida com nova letra e possíveis leves alterações melódicas. Assim temos AA. A seção B é chamada de ponte (bridge). “A ponte apresenta novo material – uma nova melodia, mudanças de tonalidades e de letra (e às vezes com maiores alterações melódicas)”<sup>46</sup> (STARR; WATERMAN, 2007, p.67). Voltaremos ao início, o A.

A música produzida pelo TIN PAN ALLEY era considerada central, *mainstream*, voltada para um público adulto, branco, da classe média para cima. Fora do eixo central, havia também a produção musical dos brancos sulistas, *hillbilly* – que originou a música *folk*, de protesto (como “Woody” Guthrie<sup>47</sup>) – e dos negros, (como Louis Armstrong e Bessie Smith). A música negra, ou “*race*”, como era também chamada, mais especificamente o *blues* e o *country blues*, trazia uma batida e uma mistura entre misticismo e sexualidade que encantou os brancos, ajudando a perpetuar obras de artistas como Charley Patton, com *Tom Rushen blues* (1929); Blind Lemon Jefferson, com *That black snake moan* (1926) e Robert Johnson com seu *Cross Road Blues* (de 1936).

<sup>44</sup> No original: “Publishing firms such as T.B Harms and Witmark and Sons hired teams of composers and lyricists to crank out new songs”.

<sup>45</sup> Originário da cidade de New Orleans, na Louisiana, advém de uma mistura de tradições musicais: ragtime, marchinhas, ritmos do *Mardi Gras* (carnaval de New Orleans), músicas francesa, italiana, caribenha, mexicana, africana e tradições espiritualistas e seculares (blues); carregando consigo diversos sentidos, entre os quais aceleração e intensificação musical; e associações sexuais. (STARR; WATERMAN, 2007, p.52).

<sup>46</sup> Do original: “the basic pattern of the lyric, and a set of chore canes to support them (...) The bridge presents new material – a new melody, chord changes, and lyrics / Finally, the A melody and chord changes are repeated with the new lyrics (and sometimes with further melodic alterations (...))”

<sup>47</sup> Woodrow Wilson “Woody” Guthrie, especialmente com as canções *This land is your land*; *Talking dust bowl blues*; *Ludlow Massacre* (1930s); *This machine Kills facists* (1940).

As músicas que integram nosso corpus de estudo são de períodos posteriores ao mencionados, todavia é importante lembrar que a música popular de hoje é fruto da influência de estilos que a antecederam, como nos informam Starr e Waterman:

*De fato, nenhum outro artista country teve maior influência sobre as gerações posteriores de músicos de blues e rock do que Johnson. Seu trabalho foi especialmente reverenciado pelo guitarrista britânico Keith Richards, dos Rolling Stones, e por Eric Clapton, cuja banda Cream lançou uma famosa cover de Cross Road Blues, de Johnson, em 1968*<sup>48</sup>

(STARR; WATERMAN, 2007, p. 106)

O período entre 1935 e 1945 foi chamado de “Era do Swing”<sup>49</sup>, trata-se, basicamente de um ritmo tocado por grandes bandas e orquestras, como Edward Kennedy “Duke” Ellington, com *Ko – Ko (1940)*; *I let a song out of my heart (1938)* e Glenn Miller, com *In the mood (1939)*, entre outras.

Já o período pós-guerra, entre 1946 e 1954, marcou uma época de mudança no panorama musical americano. Pela primeira vez os adolescentes estavam trabalhando fora de casa e ganhando seu próprio dinheiro e, conseqüentemente, passaram a consumir música<sup>50</sup> e filmes. Assim, a música que até essa época era produzida visando exclusivamente um público mais adulto, passou a enfocar os desejos dos mais jovens, abrindo caminho para o futuro *rock’n’roll*.

Apesar da mudança de público, a maior parte das músicas produzidas ainda visava ao público adulto, com intérpretes (*crooners*), que cantavam canções românticas ao som de orquestra. Por isso o gênero musical também ficou conhecido como *pop vocal*, ou seja, era a música popular da época, que privilegiava a qualidade vocal do intérprete.

Os cantores de mais destaque desse estilo foram: Frank (Francis Albert) Sinatra, que gravou mais de mil canções ao longo de sua carreira, entre elas *Nancy (with the*

<sup>48</sup> Do original: “ Indeed, no country artist had a greater influence on later generations of blues and rock musicians than Johnson. His work was especially revered by British guitarist Keith Richards of the Rolling Stones, and by Eric Clapton, whose band Cream released a celebrated cover of Johnson’s “Cross Road Blues” in 1968”

<sup>49</sup> “Swing” deriva , como jazz, do inglês afro-americano e foi usado para descrever uma fusão rítmica que levava as pessoas a se balançarem ao som de “música bem tocada” que oferecia “uma sensação de liberdade, vitalidade e diversão” (Do original: “well-played music, the term was used by extension to refer to an emotional state characterized by a sense of freedom, vitality, and enjoyment.” (STARR; WATERMAN, 2007, p. 120)

<sup>50</sup> Segundo pesquisa realizada em 1949, um terço das pessoas que compraram discos naquele ano tinham menos de 21 anos (cf. STARR: WATERMAN, 2007)

*laughing face*); *New York New York* e *Under my skin*; e Nat “King” Cole (*I love you, for sentimental reasons*, de 1946; *Nature Boy*, de 1948; *Unforgettable*, de 1950; *Mona Lisa*, de 1950 e *Too young*, de 1951).

Nas ilhas britânicas o processo não foi muito diferente. Os maiores sucessos pertenciam a músicas ‘importadas’ da América, com artistas como Bing Crosby, Rudee Vallee and ‘Whispering’ Jack Smith. Mas os ingleses também possuíam produção própria, como as bandas de música para dançar lideradas por Ambrose, Roy Fox e Jack Harris.

Nessa época um radialista de Omaha, Nebraska, percebeu que nos bares as pessoas selecionavam várias vezes a mesma música nas *jukeboxes* e, considerando o gosto do público, desenvolveu a ideia dos programas TOP 40. Em outras palavras, as quarenta canções mais pedidas e mais tocadas.

*A ideia espalhou rapidamente e, dentro de poucos anos, muitas estações de rádio estavam tocando o mesmo conjunto de músicas. A habilidade das estações de rádio em controlar a exposição do público às novas gravações levou a uma prática conhecida como **payola** [jabá], quando as gravadoras pagavam aos DJs para colocarem seus discos em ‘alta rotação’<sup>51</sup>*

(STARR; WATERMAN, 2007, p. 153)

Em 1940, os britânicos também viram o surgimento de seu primeiro ‘*Hit Parade*’ (parada de sucessos), seguindo o modelo americano e dominado por artistas americanos.

Já na periferia musical dos anos quarenta a produção se concentrava em torno da música *folk* (The Weevers) e da música sulista como o R&B<sup>52</sup> (com Louis Jordan, Muddy Waters e John Lee Hooker, entre outros) e o *country* – ou *country and western*, como era chamado na época (com Patti Page e Hank Williams).

Nos anos 50, a América testemunhou o nascimento do o *rock’n’roll*, trazendo ainda maiores mudanças para a música popular não só americana, mas mundial. A

<sup>51</sup> Do original: “The idea spread quickly, and within a few years many radio stations were playing the same set of songs. The ability of radio stations to control public’s exposure to new recordings led to a practice called **payola**, in which record companies paid DJs to put their records into “heavy rotation” [destaques do original].

<sup>52</sup> *Rhythm and Blues* – termo que teve origem nos anos 40 para descrever uma síntese dos gêneros de música negra – *gospel*, *big-band swing*, *blues* – que, juntamente com a nova tecnologia, especialmente a popularização do baixo (e da guitarra) elétrico, produziu um tipo de música popular espirituosa e propulsora. Na década seguinte seria chamado de *rock’n’roll* para camuflar as raízes negras e subsequentemente de *soul*, *funk*, *disco* e *rap*. (GEORGE, 1988, p.X; SHUKER, 2005, p. 230).



mudança mais marcante foi a abertura do mercado *mainstream* (central) para estilos musicais antes considerados periféricos ou marginais, de modo a, eventualmente, passarem a dominá-lo. Estilos como *Rhythm and blues* (R&B) e *country* passaram a ser tocados em rádios populares de sucesso central e seus discos vendidos não apenas em lojas não especializadas.

Stuessy (2008) nos lembra de que apesar da extrema importância social do *rock'n'roll* (impulsionou a economia da indústria musical, influenciou a música que ouvimos hoje, a cultura, a política, etc), trata-se primariamente de um gênero musical musical básico, porque possui “os mesmos elementos musicais presentes em outros estilos (ritmo, compasso, dinâmica, forma, timbre, etc); como, em quais proporções e com quais características esses elementos se combinam é que definem o *rock* como um estilo musical distinto”<sup>53</sup> (STUESSY, 2008, p. 1)

As letras das músicas no novo estilo abordavam temas do interesse da nova faixa do mercado, a adolescente, tais como escola; férias (em canções como *School day* e *Summertime blues*); moda ( em canções como *Black denim trousers and motorcycle boots* ou *Itsy bitsy teenie weenie yellow polka dot bikini*); danças sociais e paquera, além de se atenderem aos anseios da juventude por uma mudança musical estética.

Elvis Presley<sup>54</sup>, com canções como *Hound dog* e *Don't be cruel* (1953), foi o artista que mais colaborou para a difusão do estilo *rock'n'roll*, especialmente por sua representatividade no cinema, porém não foi o único. Entre outros artistas que também representaram o gênero na época estavam Buddy Holly (Charles Hardin Holley), com *That'll be the day* e *Peggy Sue* (1957); The Kingston trio, com *Tom Dula* (1958); Charles Edward Anderson (Chuck) Berry, com canções como *Maybellene* (1955), *Roll over Beethoven* (1956), *School day* (1957) e *Johnny be good* (1958); Richard Wayne Penniman (“Little Richard”), com canções como *Tutti-frutti*, *Long tall Sally*, *Don't*

---

<sup>53</sup> Do original” “rock and roll is basically and primarily a musical style. It utilizes the same musical elements (rhythm, pitch, dynamics, form, timbre, etc) found in all styles of music. How these ingredients are combined in what proportions and with what characteristics, defines rock as a distinct musical style.”

<sup>54</sup> Originalmente, Elvis foi contratado pela gravadora RCA como um cantor de Rockabilly (um rock caipira, ou seja, uma mistura de *rock'n'roll* com o estilo *hillbilly* sulista), mas foi transformado na primeira estrela de *rock'n'roll* (STARR; WATERMAN, 2007, p.209).

*knock the rock*, *The girl can't help it* e *Mister rock'n'roll* (gravadas entre 1956 e 1957); e Antoine “Fats” Domino, com *I'm in love again* e *I'm walking*.

O *rock'n'roll* e toda a sua movimentação corporal estimularam o surgimento de um estilo dançante que predominou o início da década de 1960, o twist (e.g. Chubby Checker e as Ronettes) e colaborou para que artistas negros passassem a conquistar, ainda com muita dificuldade, sua faixa de mercado – graças à criação da gravadora periférica Motown, em Detroit, Michigan, que em meados dos anos sessenta levou ao estrelato artistas como The Supremes (de onde saiu a cantora solo Diana Ross), especialmente com a canção *You can't hurry love*, de 1966; e The temptations, com *My girl*, em 1965.

Os anos sessenta foram marcados pela abertura musical para novidades estilísticas, no que diz respeito à música popular, mais especificamente a América foi invadida pela música advinda de outras terras, especialmente das ilhas britânicas.

*Embora a América tenha sempre tido uma fascinação cultural pelas coisas britânicas ao longo da história, a Beatlemania americana representa de fato a primeira vez que esse nível de adulação foi garantido a músicos pop não nativos. A América exportou sua música popular para a Grã Bretanha, para a Europa e cada vez mais para o mundo industrializado com enorme sucesso ao longo do tempo, mas o impacto dos Beatles neste país marcou o significativo início de um processo recíproco e agressivo.*<sup>55</sup>

(STARR; WATERMAN, 2007, p. 254)

Como podemos observar, a invasão dos Beatles abriu as portas para o início do processo de importação de música popular estrangeira, especialmente a britânica. Esse processo não foi muito doloroso, uma vez que, como já vimos, a música popular americana foi construída através de um processo de junções de estilos advindos de variadas fontes culturais e, por outro lado, os Beatles, assim como outras bandas britânicas, também haviam sido influenciados pela música americana anteriormente.

Embora tenham suas raízes musicais atreladas ao *rock'n'roll*, os Beatles são considerados mais frequentemente como uma banda do estilo *pop-rock*, pois englobam as características do *rock'n'roll* associadas à sua popularidade. Seu refinamento musical atrelado ao abandono do vocal vigoroso lhes confere lugar no gênero *pop* (cf.

<sup>55</sup> Do original: “Although America has retained a cultural fascination with things British throughout its history, American *Beatlemania* does represent the first time this degree of adulation was bestowed on non-native pop musicians. America had been exporting its popular music to Great Britain, to Europe, and increasingly throughout the industrialized world with enormous success for a long time, but the impact of the Beatles in this country marked the significant beginning of an aggressively reciprocal process.”

CHARLTON, 2009; Rock and Roll Hall of Fame, 1988<sup>56</sup>; The Rolling Stone Magazine; all-music<sup>57</sup>). Destacamos, entre os sucessos do Beatles nos EUA as canções *Please, please me* e *I wanna hold your hand*, de 1963; *Norwegian wood*, de 1965; e *Sgt. Pepper's lonely hearts club band*, de 1967.

Outras bandas seguiram o caminho aberto pelos Beatles, entre elas The Rolling Stones, especialmente com *As tears go by*, de 1966 e *Harlem Shuffle*, de 1964; The Animals; The Who; The Kinks; Cream (de onde saiu Eric Clapton), entre outros, que foram categorizados como estilo simplesmente *rock*. *Rock* é, segundo Charlton (2009, p.183-186), um gênero de música popular marcado pelo som de guitarra e baixo elétricos, teclado e bateria, utilizando-se normalmente de três acordes, melodia agradável e contratempo.

Outros estilos musicais, já estabelecidos no país, continuaram suas produções. A música *country* desse período foi representada, entre outros por Johnny Cash, com *Hey Porter*; *Cry, cry, cry*; e *I walk the line* e Patsy Cline, com *I fall to pieces* e *Crazy* (1961), entre outros. O estilo *country soul*<sup>58</sup>, ou simplesmente *soul*, predecessor do R&B, foi representado especialmente por Ray Charles (*What'd I say*, *Georgia on my mind* e *I've got a woman*); James Brown (*Please, please, please*; *There was a time*; *Say it loud – I'm black and I'm proud* e *Papa's got a brand new bag*) e Aretha Franklin (*I never loved a man (the way I love you)*; *Freward of love* e *Respect*).

A costa oeste contribuiu com o surgimento de um *rock* “praiano” (*surf rock*), que segue as características musicais do *rock'n'roll*, trazendo em suas letras a temática da vida dos surfistas, das pessoas que moram na costa oeste americana, suas preocupações ecológicas e seu misticismo. Seu principal representante é a banda Beach Boys, inicialmente chamada Brian Wilson and the Beach Boys, com canções como *Surfin USA* (1963), *California girls* e *Good vibrations* (1966).

O período também teve representatividade no estilo *folk*, ou canções de protesto, com artistas como Joan Baez, Judy Collin e, principalmente, Bob Dylan (Robert

<sup>56</sup> Conforme o site: <http://www.rockhall.com/inductee/the-beatles> (consulta 10/12/2009).

<sup>57</sup> allmusic.com – site especializado em musicologia (consulta 27/08/2009).

<sup>58</sup> Estilo *country* acrescido de raízes negras, com movimento e improvisações, cheio de sentimentalismo e força musical, geralmente acompanhado de piano (Charlton, 2008, p. 87); com formato de três versos e duas estrofes (bar), tocada com realismo social, com tons de beleza, tortura, angústia e desespero nos vocais (SHUKER, 2005, p.25).

Zimmerman), com canções como *Where have all the flowers gone* (1962); *Blowing in the wind* (1963) e *Like a rolling stone* (1965).

O envolvimento da música popular com manifestações populares de cunho sócio-político do final dos anos sessenta fez parte de um movimento que foi chamado de contra-cultura, sendo representado por músicos como The Grateful Dead; Jefferson Airplane; Janis Joplin; Jimmy Hendrix; Eric Clapton e Credence Clearwater Revival.

Como vimos, a música dos anos sessenta apresentou uma grande variedade e riqueza de estilos. O Twist, que foi a música dançante da época, deu lugar, nos anos de 1970 à *disco music*, que era um estilo caracterizados por elaboradas gravações de estúdio e foi fortemente representado por artistas como Donna Summer, Chic, Village People e Bee Gees.

Na realidade, segundo Starr e Waterman (2007, p.317) a parada dos sucesso dos anos setenta, representada pela *Billboard charts*, mostra a presença de diversos estilos/gêneros de música popular como Tin Pan Alley; *soul*; *country* e *rock*. As rádios foram criando novos títulos a fim de classificarem novas músicas derivadas do rock, entre as quais podemos apontar as seguintes representações<sup>59</sup>: o *pop rock* (Peter Frampton); a música adulta contemporânea (Barbara Streisend); o cantor- compositor (Bob Dylan; James Taylor); *soft soul* (Barry White); *country pop* (John Denver); *bubblegum* (Jackson five); e *disco* (Donna Summer, Bee Gees).

O *top-40* (as quarenta músicas mais tocadas) da época apresentava também artistas como Steve Wonder; Carole King; Paul Simon; Frank Zappa e Neil Diamond, entre os “estrangeiros” Led Zeppelin; (*rock*) Elton John (*pop*); Pink Floyd (progressivo<sup>60</sup>); Crosby, Stills and Nash; Rolling Stones; Santana e David Bowie. (p.329)

Esse amálgama de estilos fez com que a característica predominante da nova música *rock* fosse a “experiência imediata da textura musical, do momento rítmico e da intensidade emocional” (STARR; WATERMAN, 2007, p. 332). Diferente da música do início do século, muitas músicas-*rock* não apresentavam em suas letras uma história

<sup>59</sup> adaptado das paginas 314 e 315 de STARR; WATERMAN, 2007.

<sup>60</sup> Estilo musical que se originou do rock, com influencias R&B, mas com melodias complexas, músicas longas (podendo untrapassar 15 minutos) , harmonias e escalas musicais complexas com compassos compostos, contendo a presença de instrumentos musicais alternativos ao rock como violino e flauta ou teclado sintetizado. (CHARLTON, 2009, p. 227-228).

com estrutura lógica ou uma imagem organizada , mas sim “encontravam sua unidade na coesão visceral do estilo musical com textura e atitude”<sup>61</sup> (Ibid.)

Apesar de todas essas características e mudanças, o mercado de música popular ainda considerava algumas representações como mercado periférico. Entre eles estavam *country progressivo* (Willie Nelson; Townes Van Zandt); *reggae*, (Bob Marley; Johnny Nash; Jimmy Cliff); *punk* (Ramones; Who; Patti Smith; Sex Pistols; Clash; Damned; Blondie e Talking Heads); *funk* (Sly Stone e George Clinton); e o *rap*, precursor do *Hip Hop* (Grandmaster flash e Afrika Bambaataa).

Precisamos destacar o *punk*, pois dos gêneros que se originaram dele, dois fazem parte de nosso corpus de estudo: *punk rock*<sup>62</sup> e *punk pop*<sup>63</sup>. O *punk* não surgiu como um gênero musical, mas como um movimento sociopolítico, fruto de uma juventude descontente e desencantada que agressiva e violentamente procurava conquistar seus ideais anarquistas, niilistas e políticos, especialmente no que diz respeito à liberdade de expressão. Musicalmente, o estilo é barulhento, rápido, agressivo, resumindo-se a três acordes, tocados com muita atitude no baixo, guitarra e bateria (BENNETT, 2007).

Pudemos perceber até o momento que, embora tenha recebido diversas influências e aceitado muitas mudanças, o mercado da música popular ainda continuou, de certa forma, sob o domínio da indústria musical central. Esse domínio aumentou com a digitalização das gravações, mas, conseqüentemente também proporcionou o acesso a mais variações musicais nos anos oitenta.

Nos anos de 1980, a tecnologia digital e o advento da MTV (em 1981) marcaram novas mudanças no mercado musical popular *mainstream*. Tecnicamente, a mudança de sistema de gravação do analógico para o digital trouxe os sintetizadores (aparelhos que sintetizam sons musicais), as baterias eletrônicas, o CD e a MIDI<sup>64</sup>.

---

<sup>61</sup> Do original: “the immediate experience of musical texture, rhythmic momentum, and emotional intensity (...)but rather find their unity in a visceral cohesion of musical style, texture and attitude.”

<sup>62</sup> *Punk rock* é o rótulo utilizado a partir dos anos 90 para a versão subsequente ao *punk*, que conserva o formato musical, e predomina a atitude política. (cf. DANTAS, 2007).

<sup>63</sup> *Punk pop* é o rótulo utilizado, também a partir dos anos 90 para uma versão mais comercial/popular da música *punk* (cf. DANTAS, 2007).

<sup>64</sup> Musical Instrument Digital Interface, protocolo de comunicação entre vários instrumentos, que permite gravar e equalizar sons.

A MTV proporcionou uma maior fusão entre imagem e som. A partir daquele momento, o artista que idealizasse ter sucesso deveria, além de gravar a música, apresentar um vídeo musical que agradasse seu público.

Artistas que conquistaram seu espaço na música popular dos anos oitenta são, entre outros: Kenny Rogers, com *Lady* (1980); Eurythmics, com *Sweet dreams* (1983); Tina Turner, com *What's love got to do with it* (1984); Van Halen, com *Jump* (1984); Peter Gabriel, com *Sledgehammer* (1986); Michael Jackson, com o álbum *Thriller* (1983), que foi o álbum número um na parada de sucessos americana por 37 semanas (p. 397), com as músicas *Thriller*; *Billie Jean* e *The girl is mine*; Bruce Springsteen, com *Born in the USA* (1984); Paul Simon, com *Graceland* (1986); Prince, com *Purple rain* e *When doves cry* (1984); e Madonna, com *Like a virgin* (1984).

O sucesso de artistas *pop* como Madonna e Michael Jackson está claramente ligado à indústria do vídeo, consequentemente à MTV (MUNDY, 1999). Esses artistas deixaram-se explorar como uma mercadoria visual, lançando modismos e danças. Nos anos oitenta, ao redor do mundo, meninas começaram a deixar as alças do sutiã para fora das blusas e a usarem luvas de renda sem as pontas dos dedos. Os meninos todos queriam uma jaqueta alaranjada, como a de Jackson. E todo mundo queria aprender a fazer o “*moonwalk*” (dança que simula um deslizar para trás, criado por Jackson). Em suma, essa relação entre música popular e consumo marcou a presença da música no discurso de ideologia e entretenimento, rotulando esses artistas como *pop* (CHARLTON, 2009, p. 306; 333).

Naquele momento também a música deixou de focalizar um público pertencente a uma faixa etária exclusiva, como nos períodos anteriores. A união de imagem (cinema, televisão, propagandas) e música popular criou um novo estilo de vida que Mundy (1999, p. 232) chama de “jovens de todas as idades”<sup>65</sup>, pois o jovem da geração pós-guerra, cheio de conflitos históricos, para quem a música *rock* representava oposição política, diferença e autenticidade, deixou de existir.

Mundy ainda nos lembra que, embora a relação entre *rock* e TV tenha sido contraditória numa época, nos anos 80 ela deixou de ser:

---

<sup>65</sup> Do original: “young people of all ages”

*A habilidade da televisão em mediar a música popular para a construção da comunicação global ficou aparente no concerto Live Aid, de 1985, no qual a TV ligou o estádio de Wembley, em Londres, ao estádio JFK, na Filadélfia, assim como o show [para a libertação] de Nelson Mandela, em 1988<sup>66</sup>*

(MUNDY, 1999, p. 213-214)

O que Mundy quer salientar é que, no século vinte, mais especialmente a partir da década oitenta, o sucesso de músicas populares se deu na maioria das vezes em virtude de sua união à imagem, pois o vídeo musical tornou-se fator dominante na comercialização musical. Em outras palavras, os vídeos que não eram “selecionados” como parte do *mainstream*, ou os artistas que não gravavam vídeos de suas canções, continuariam a percorrer o caminho periférico.

Vimos até agora, portanto, que, dadas as devidas diferenças, na música popular americana esteve sempre presente uma divisão entre música central e periférica. Se no começo do século havia uma clara divisão entre as canções românticas do Tin Pan Alley e as músicas jazz para dançar como centrais, em oposição à música “*hillbilly*” e “*race*”, quanto mais nos aproximamos do final do século, menos essa divisão vai ficando tão clara.

De acordo com Starr e Waterman (2007, p.420), isso acontece porque a música popular que vende mais não apresenta marcas estilísticas tão diferentes umas das outras. Segundo eles, se considerarmos os álbuns mais vendidos nos anos de 1990, por exemplo, a variedade é ainda maior que na década anterior. Entre os artistas que mais venderam nos anos noventa estão<sup>67</sup>: (as divas *pop vocal*) Celine Dione, Janet Jackson e Mariah Carey – a artista pop e R&B que mais vendeu na década; (*country*) Clint Black, Reba McEntire, Shania Twain, e Garth Brooks – o artista masculino que mais vendeu na década, não só no *country*; (*R&B/rap*) Boyz II Men, Soop Doggy Dogg, 2Pac Shakur, e the Notorius B.I.G; (*hard rock*)<sup>68</sup> Aerosmith; (cantora-compositora alternativa)

<sup>66</sup> Do original: “the ability of television and its mediation of popular music to construct imagined (p.214) global communication was apparent in the 1985 Live Aid concert, in which TV linked Wembley Stadium in London and the JFK Stadium in Philadelphia, and in the 1988 Nelson Mandela concert”.

<sup>67</sup> Adaptado de STARR; WATERMAN, 2007, p.420-421.

<sup>68</sup> O *hard rock* é um estilo “poderoso e agressivo que se desenvolveu a partir do blues dos anos sessenta”, é antecessor do *heavy metal*, apresentando menor intensidade, com forte associação à sexualidade e masculinidade do vocalista, com melodia crua e forte bateria. (CHARLTON, 2009, p. 198; SHUKER, 2005, p. 130-131).

Alanis Morissette; (*heavy metal*)<sup>69</sup> (Metallica); além das trilhas sonoras dos filmes *The bodyguard* (1993) e *Titanic* (1996).

Essa variedade denota o certo desespero das gravadoras em tentarem replicar o sucesso do fenômeno Michael Jackson, investindo milhões de dólares em diversos estilos diferentes. Essa prática também deu origem a um outro fenômeno musical chamado de “one-hit wonders”, ou seja, artistas que ficaram extremamente famosos por uma única canção e nunca mais conseguiram atingir sucesso. Nos anos de 1980, podemos citar os ingleses do Kajagoogoo, com *Too shy* e do Frankie goes do Hollywood, com *Relax*; a alemã Nena, com *99 Luftballons*; os australianos do Midnight Oil, com *Beds are Burning*; e o americano Bobby McFerrin, com *Don't Worry Be Happy*. Os anos de 1990 estão representados, entre outros, por Joan Osbourne, com *One Of Us*; 4 Non Blondes, com *What's Up*; Right Said Fred, com *I'm Too Sexy*; Blind Melon, com *No Rain* e Los Del Rio, com *Macarena*<sup>70</sup>.

Podemos perceber, portanto, que a música popular dos anos 90, assim como a da primeira década do ano 2000, apresenta os mais variados rótulos para novos artistas cuja música apresentou-se influenciada pelo R&B, pelo *rock*, pelo *country*, etc.: *Hip Hop*; música *Techno* (para dançar); *pop/rock* alternativo; *grunge*<sup>71</sup>; *country* alternativo; *country rock* alternativo; música cristã contemporânea alternativa; metal alternativo; *rap* alternativo; *pop/rock* alternativo<sup>72</sup>; cantor-compositor alternativo; *world music* (música com influência de outras culturas, especialmente a africana).

O esboço que procuramos traçar sobre a história da música popular em língua inglesa teve como objetivo oferecer ao leitor deste trabalho uma visão do processo de transformação musical que culminou em tantos rótulos, alguns dos quais foram selecionados para nossa análise, como apresentaremos no capítulo metodológico.

---

<sup>69</sup> *heavy metal* estilo “baseado na música pesada, distorcida e nos signos imagéticos utilizados pelos *headbangers* (cabelos compridos, calças jeans “surradas” e camisetas de bandas (...)) [com] ritmo binário, marcado forte” (JANOTTI JR, 1994, p.10); predominantemente orientado pela guitarra, utilizando-se de apelo para a massa e energia musical (SHUKER, 2005, p.133).

<sup>70</sup> Segundo o site <http://www.onehitwondercentral.com>

<sup>71</sup> Estilo alternativo dos anos noventa que musicalmente apresentava uma junção dos efeitos distorcidos das guitarras do final dos anos 60 e do início dos anos 70, numa fusão de estilos (*heavy metal*, *hard rock*, etc.) associados a um vocal rouco que interpreta letras melancólicas, sarcásticas e críticas, cantadas (muitas vezes faladas) por vocais roucos e profundos (cf. SHUKER, 2005, p.129).

<sup>72</sup> O termo alternativo foi adotado para marcar uma variedade do estilo principal, especialmente marcada pela fusão desse com outros estilos, seja música que denote a origem do país ou a combinação de vários outros estilos de música. (CHARLTON, 2009, p. 344).



Encerramos o capítulo ressaltando o aspecto de abrangência e agregação da música popular. Se a música é considerada de boa ou de má qualidade não cabe a nós, enquanto pesquisadores, definir. Cabe-nos, contudo, observar a habilidade da música de “ultrapassar as barreiras da sociedade e os limites das nações, [de manter] aberta a possibilidade de que podemos ver algo familiar e simpático em pessoas estranhas a nós”<sup>73</sup>. Ao nos disponibilizarmos a aceitar a música popular do outro também como parte da nossa, estamos reconhecendo que a música popular nos oferece um elo entre aquilo que somos, ou fomos, em nossa adolescência para aquilo que passamos a ser, ou seremos.

O capítulo seguinte apresenta os preceitos teóricos linguísticos que nortearam a nossa análise.

---

<sup>73</sup> Do original: “Music, with its ability to flow over the boundaries of society and the borders of nations, holds open the possibility that we may glimpse something familiar and sympathetic in people strange to us”

## **2. Fundamentação Teórica**

Neste capítulo discutiremos os fundamentos que embasaram nossa pesquisa teórica e metodologicamente. Primeiramente apresentaremos a teoria que serviu como sustentáculo teórico para nosso trabalho que é a Linguística de Corpus, discutindo os conceitos de corpus e representatividade, a visão de língua por trás desses conceitos, a influência de contexto para a análise e as diferentes nuances de texto.

Em seguida, apresentamos os pressupostos metodológicos da Análise Multidimensional, conforme propostos por Biber (1988) que nortearam nossa análise. Descrevemos os passos da pesquisa de Biber, as dimensões de análise textual por ele desvendadas para a língua inglesa e resenhamos dois trabalhos brasileiros que seguiram esse modelo.

Encerramos o capítulo apresentando uma breve discussão sobre o conceito de gênero que adotamos para este estudo, considerando sua relação com a visão de linguagem subjacente à Linguística de Corpus e sua interdependência com pesquisas de análise multidimensional.

### **2.1. Linguística de Corpus**

Conforme dito no capítulo introdutório, o trabalho aqui descrito (tanto o estudo piloto como o principal) tem como fundamentação teórica principal a Linguística de Corpus (LC). A LC é uma área que estuda os fenômenos da língua por meio da observação de grandes quantidades de dados empíricos armazenados eletronicamente, por computador. Não se pode pensar, contudo, que se trata simplesmente da análise de coleções de textos, essencialmente porque, a coleta dos corpora é feita de modo criterioso e com propósito claro e bem definido. Apresentaremos, nesta seção pontos teóricos importantes para a LC e concepções que a incorporam, partiremos da definição de corpus, suas características e representatividade, para o conceito de linguagem e

análise linguística (que envolvem a questão do gênero/registo) que estão atrelados à própria definição de corpus e os desdobramentos do corpus que sustentam as pesquisas em LC.

Entre as diversas definições para corpus apresentadas por estudiosos, consideramos, assim como Berber Sardinha (2004a), a proposta de Sanches (1995, p. 8-9):

*“Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise”*

(BERBER SARDINHA, 2004a, p.18)

Trata-se da definição mais completa por mencionar vários pontos importantes, a saber: a origem (dados autênticos); o propósito (“deve ser objeto de estudo linguístico”); a composição (escolha criteriosa); a formatação (dados “legíveis por computador”); a representatividade (“o corpus deve ser representativo de uma língua ou variedade); e a extensão (“deve ser vasto para ser representativo”) (BERBER SARDINHA, 2004a, p. 18-19). Segundo Berber Sardinha, as quatro características fundamentais para a coleta dos dados formadores de um corpus computadorizado são textos em linguagem natural, autenticidade, critério rigoroso e representatividade. Ao se observar tais critérios, cuidar-se-á para que os textos advenham de ocorrência natural, não sejam produzidos especialmente para o fim da pesquisa nem em linguagem artificial (“tal como linguagem de programação de computadores ou notação matemática”) (id.), que sejam produzidos por falantes nativos e “quando esse não é o caso, deve-se qualificá-lo como corpora de aprendizes” (id.); que sejam coletados de forma que objetivem responder as perguntas dos pesquisadores seguindo critérios por esses pré-estabelecidos; e representem de fato uma amostra da população que se pretende analisar.

A questão da representatividade é um pouco mais delicada, pois embora, ainda segundo Berber Sardinha refira-se “justamente à extensão do corpus”; a visão probabilística da língua subjaz a necessidade de grande extensão, pois “quanto maior a

quantidade de palavras, maior a probabilidade de aparecerem palavras de baixa frequência” (BERBER SARDINHA, 2004a, p. 22-23). Por outro lado, o autor observa que se devem considerar as questões: “*representativo do quê?*” e “*representativo para quem?*”. Isso implica em coletar a maior quantidade de palavras possíveis, a maior quantidade de textos e dos mais variados gêneros, registros ou tipos (id., *ibid.*, p. 25). Sob essa perspectiva, o autor sugere que “o tamanho mínimo necessário para um corpus de estudo depende (a) do que se pretende estudar e (b) da especialização do corpus” (id., *ibid.*, p. 98).

Ainda assim, o autor sugere a seguinte classificação para a extensão de um corpus, “baseada na observação dos corpora utilizados, segundo quatro anos de conferências de Linguística de Corpus” (BERBER SARDINHA, 2004a, p. 26):

<b>Tamanho em palavras</b>	<b>Classificação</b>
Menos de 80 mil	Pequeno
80 a 250 mil	Pequeno-médio
250 mil a 1 milhão	Médio
1 milhão a 10 milhões	Médio-grande
10 milhões ou mais	Grande

Quadro 1: Classificação de corpora por número de palavras. Reproduzido de Berber Sardinha (2004a, p. 26)

O corpus coletado para esta pesquisa encaixa-se nas características fundamentais apontadas por Berber Sardinha. Os textos por nós coletados advêm de fonte natural, pois foram produzidos por falantes nativos para outros fins que não o da pesquisa; são autênticos, pois estiveram sujeitos à interação natural com o leitor/ouvinte (são letras de músicas realmente gravadas e veiculadas pela mídia); e são representativos de letras de música, vez que correspondem a uma amostra significativa desse universo linguístico, mais de 1 milhão de palavras<sup>74</sup>, sendo, portanto, classificado como um corpus médio-grande. Nosso corpus de estudo será detalhado no capítulo metodológico.

Questões como naturalidade, autenticidade e representatividade que envolvem a montagem de um corpus estão vinculadas à visão de linguagem adotada pelos linguistas

<sup>74</sup> Biber (1988) utilizou um corpus com 960 mil palavras como amostra para traçar diferenças entre a língua escrita e a falada como um todo (cf. será discutido na seção posterior)

do corpus, que a consideram como um sistema probabilístico. Esta acepção de linguagem está enraizada na observação empírica<sup>75</sup> da língua, ou seja, na observação natural da ocorrência de dados linguísticos em situações reais de uso, a fim de explicar os fenômenos linguísticos em sua função social. Não se trata de uma definição que teve origem com a LC, mas advém principalmente de Halliday e Sinclair, seguindo as ideias de Firth (FIRTH, 1957; HALLIDAY; HASAN, 1989; HALLIDAY; 1991; SINCLAIR; 1991; HALLIDAY; WEBSTER; 2002).

A visão probabilística assume que, embora muitas escolhas e combinações lexicais sejam possíveis, dentro de um determinado contexto (e na língua em geral), elas não ocorrem da mesma forma nem com a mesma frequência (BERBER SARDINHA, 2004a, p. 30). A variação de escolhas linguísticas não ocorre aleatoriamente, ela segue uma certa padronização representativa de cada gênero, como aponta o autor:

*“o conjunto da pesquisa desenvolvida por Biber apresenta evidências inequívocas de que conjuntos de traços linguísticos variam sistematicamente com relação a textos típicos de contextos comunicativos específicos. Em outras palavras, a variação não é aleatória.”*

(BERBER SARDINHA, 2004a, p. 31)

Em outras palavras, as combinações lexicais, ou padrões, têm maior ou menor probabilidade de ocorrerem, ou são mais ou menos frequentes, de acordo com o contexto. A observação empírica vai permitir uma descrição mais completa e mais precisa da língua, pois evidenciará aquilo que é mais típico (HALLIDAY; WEBSTER, 2002, p. 10) em um determinado contexto.

Assim a língua pode ser observada segundo seus aspectos funcionais (as funções que os falantes querem desempenhar, aquilo que querem dizer) e seus aspectos formais (as escolhas estruturais do falante), por isso, a abordagem analítica proposta por Halliday é chamada de sistêmico-funcional. Em suma, a língua, enquanto sistema, oferece possibilidades de escolhas de padrões léxico-gramaticais dentro de um contexto social e situacional, o qual determina essas mesmas escolhas de acordo com o significado, ou a função que se quer desempenhar. Segundo Halliday e Hasan (1989, p. 24-30), as funções

---

<sup>75</sup> “as afirmações feitas acerca da língua devem ser fundamentadas pela evidência do uso da linguagem” (KENNEDY, 1998, p.8)

se manifestam segundo três grandes categorias (ou metafunções) de análise de significado (semânticas): ideacional (campo do conteúdo, ou *field*); interpessoal (interação entre os participantes, ou *tenor*) e textual (organização do texto, ou *mode*).

Já a metodologia de análise multidimensional utilizada nesta pesquisa, conforme proposta por Biber (1988; BIBER; CONRAD, 2009), também assume as características semântica e funcional da língua, pois admite que “*textos podem ser descritos de acordo com seus contextos, considerando as características das pessoas que produziram os textos e as características das situações e os propósitos comunicativos associados ao texto*”<sup>76</sup> (BIBER; CONRAD, 2009, p.5). Em outras palavras, a análise das características linguísticas em um texto deve considerar o contexto de situação e vice-versa.

O modelo de análise multidimensional requer que o contexto situacional de comunicação seja levado em consideração na interpretação das dimensões<sup>77</sup> geradas a partir dos fatores<sup>78</sup> extraídos pela análise estatística. Por outro lado, o modelo para AMD de Biber, diferente de outros modelos de análise linguística, que observam a língua de acordo com as funções, considera primeiramente as evidências, ou seja, os agrupamentos de características recorrentes, para, em seguida, focar as funções por elas desempenhadas. Isso possibilita a “descoberta de novas funções criadas pelos usuários da língua em diferentes situações e com diferentes propósitos” (PACHECO DE OLIVEIRA, 1997, p.19). Esse modelo será descrito na próxima seção e no capítulo metodológico.

O conceito de contexto de situação encontra-se fortemente atrelado à noção de gênero e registro. Como dissemos na introdução deste trabalho, os termos gênero e registro estão longe de atingir definições consensuais e não é objetivo deste trabalho esgotar tal discussão. Adotamos a perspectiva de que tanto um como outro representam diferentes perspectivas de análise textual, sendo o primeiro (gênero) o nível onde

---

<sup>76</sup> Do original: “Texts can be described according to their contexts, considering the characteristics of the people who produced the texts, and the characteristics of the situations and communicative purposes associated with the texts”

<sup>77</sup> Dimensões correspondem a níveis diferentes de análise de texto de acordo com suas funções sócio-situacionais.

<sup>78</sup> Fatores representam os agrupamentos de variáveis (característica linguísticas) que co-ocorrem

observamos os propósitos comunicativos a partir da totalidade das estruturas de construção textual, e o segundo (o registro) o nível onde essas construções podem ser observadas a partir das características sintáticas, morfológicas e semânticas. Apresentaremos, na última seção deste capítulo, uma discussão mais detalhada sobre esta questão.

Respeitando as diferenças entre os diversos gêneros e registros e observando seus propósitos comunicativos, os estudiosos da Linguística de Corpus utilizam grandes quantidades de dados linguísticos autênticos como fonte de observação das prováveis escolhas do falante e sua possibilidade de uso na língua (entre outras possibilidades de estudo). Em outras palavras, a característica predominante das pesquisas baseadas em corpus é o uso de corpora como fonte de informação representativa da linguagem natural dos falantes (BERBER SARDINHA, 2004a, p.32).

Os corpora podem ser de natureza exploratória ou confirmatória. Esta, busca nos dados exemplos que confirmem teorias pré-existentes, e aquela observa, nos dados, padrões e diferenças que levarão a generalizações sobre a língua. Krieger (2003) ressalta, contudo, que o objetivo da LC não é o de gerar teorias, mas o de buscar evidências sobre padrões e comportamentos das palavras, identificados a partir de dados empíricos.

Pesquisas de natureza exploratória, ou o termo em inglês *corpus-driven* (movida a corpus) (TOGNINI-BONELLI, 2002; BERBER SARDINHA, 2004a) não têm a intenção de levantar exemplos, mas de fornecer novas evidências a partir da observação da língua nos corpora, assim apresentam “inclinação para desafiar a descrição linguística e oferecer novas perspectivas fascinantes sobre a língua”<sup>79</sup> (TOGNINI-BONELLI, 2002, p. 74). Entre diversas pesquisas que utilizaram abordagem de natureza exploratória podemos citar o projeto COBUILD (pesquisa lexicográfica) e os trabalhos de Hunston e Francis (2002); Sinclair (1991) e Tognini-Bonelli (2001).

Pesquisas de natureza confirmatória são pesquisas baseadas em corpus, ou, em inglês, *corpus-based*. Biber (TOGNINI-BONELLI, 2001; BIBER et al, 1998, p. 9-12)

---

<sup>79</sup> Do original: “... is bound to challenge language description and offer fascinating new insight into language.”

observa que as pesquisas baseadas em corpus devem servir como “abordagem complementar às abordagens mais tradicionais<sup>80</sup>”, pois frequentemente perguntas que levam à investigação de corpora advêm de outras áreas de investigação, tais como: análises estruturais<sup>81</sup>; análises para confirmar a frequência de certas evidências intuitivas segundo diferentes registros; investigações para detectar ‘erros’ recorrentes de alunos de segunda língua ou de língua materna; análises detalhadas de características linguísticas específicas em textos específicos, para complementar descobertas em corpora maiores; análises de interações conversacionais; etc. Enfim, o autor sugere que a abordagem baseada em corpus “pode ser utilizada em quase qualquer área de investigação linguística empírica”<sup>82</sup> (BIBER, et al, 1998, p.11), como a lexicografia (e.g., associações de palavras, colocados, frequência, etc.), a gramática; a sociolinguística (e.g., pesquisas sobre dialeto); a educação (e.g., padrões na linguagem de aprendizes, contraste entre falantes de L1 e L2, confecção de material didático); a estilística (e.g., estilos particulares de escritores ao longo do tempo); entre outras.

É preciso salientar que a natureza exploratória não necessariamente exclui a natureza confirmatória de uma pesquisa, como é o caso da pesquisa de Biber (1988), por exemplo, que inicialmente adota uma abordagem confirmatória (*corpus-based*), partindo da observação do corpus, mas que passa a utilizar conceitos e categorias preexistentes para selecionar as características linguísticas (variáveis) que o levaram ao estudo das dimensões de linguagem e tipo de texto. Pesquisas como as de Pacheco de Oliveira (1997) e de Kauffmann (2005), que apresentaremos na seção seguinte, também podem ser consideradas intermediárias, como a de Biber. O mesmo pode ser dito sobre a pesquisa que ora descrevemos, pois partimos nós também do corpus de estudo como fonte de evidência, e, em seguida, selecionamos as variáveis de acordo com critérios pré-estabelecidos, cuja análise resultou nas dimensões que expressam funções comunicativas em letras de música.

---

<sup>80</sup> Do original: “complementary approach to more traditional approaches”

<sup>81</sup> e.g., investigar o uso de orações relativas implica em saber *a priori* da existência de tal estrutura

<sup>82</sup> O autor oferece uma série de exemplos, entre eles: “pesquisas sobre palavras isoladas, características gramaticais, linguagem de homens e mulheres, aquisição de linguagem por crianças, estilo de determinado autor, padrões de registro” (BIEBR, et al, 1998, p.11)



Entre inúmeros pesquisadores da área destacamos especialmente no Brasil, o trabalho pioneiro de Tony Berber Sardinha, fundador do Grupo de Estudos de Linguística de Corpus (GELC), que vem desenvolvendo e orientando pesquisas baseadas em LC, com foco em metáfora, ensino-aprendizagem de língua estrangeira, linguística aplicada e linguística sistêmico-funcional. Entre seus principais trabalhos podemos destacar o estudo sobre gênero e padronização lexical (BERBER SARDINHA, 2003), sobre identificação de metáforas com corpora eletrônicos (BERBER SARDINHA, 2007); e, sobretudo, o desenvolvimento, com colaboradores, do corpus do Português Brasileiro, desde 2005 (BERBER SARDINHA, 2005) e do desenvolvimento de ferramentas computacionais para análise de corpora<sup>83</sup>.

Como vimos, a LC pode ser utilizada em diferentes contextos de pesquisa, estendendo, pois, nossa definição de corpus, consideramos que sua coleta está sujeita a diferentes propósitos, representatividade e formato. A fim de atender as exigências de cada pesquisa, os corpora se distinguem especialmente entre o corpus geral e o corpus especializado. O primeiro é uma coletânea de textos naturais, incluindo diferentes gêneros e domínios, textos falados e escritos, particulares ou públicos (Kennedy, 1998, p. 20), geralmente utilizados para explorar a linguagem, visando à observação de comportamentos específicos sobre vocabulário, gramática ou estrutura do discurso de uma língua como um todo. Esse tipo de corpus armazena enorme quantidade de dados a fim de ser representativo da língua geral. O segundo é desenvolvido para atender a necessidades específicas de um trabalho de pesquisa, seguindo critérios específicos da pesquisa, considerando a população, os objetivos, o gênero ou gêneros contidos no corpus, assim como a especificação do período de tempo. Esse tipo de corpus pode servir como fonte de caracterização da língua apenas em relação ao corpus de estudo específico (McEnery & Wilson, 1996, p. 64).

O corpus coletado para esta pesquisa é um corpus especializado. Além de ter sido coletado especialmente para atender aos critérios da pesquisa, assim como pode ocorrer com outros corpora especializados, é usado como fonte de estudo contrastivo. Ou seja,

---

<sup>83</sup> vide <http://www2.lael.pucsp.br/corpora/>

um corpus especializado pode ser comparado a um corpus de língua geral, chamado de corpus de referência nesse caso, possibilitando a observação e destaque das especificidades de corpus de estudo em relação a outros gêneros linguísticos presentes na língua estudada.

Hoje em dia existem corpora geral representativos de diversas línguas. No caso da língua inglesa, que é objeto do nosso estudo, podemos citar especialmente os dois que a integram: O BNC e o ANC. O British National Corpus (BNC) tem 100 milhões de palavras (cf. site institucional)<sup>84</sup> e, segundo Berber Sardinha (2004a, p. 8), “*é de destaque porque foi o primeiro a conter 100 milhões de palavras e ainda é, entre os megacorpora, o único disponível para compra*”. O American National Corpus (ANC) está sendo desenvolvido seguindo os moldes do BNC para se tornar um corpus comparável. Até o momento, esse corpus conta com 22 milhões de palavras disponíveis.

Embora não integrantes de nossa pesquisa, é preciso ressaltar outros corpora de inglês geral por sua representatividade. Primeiramente, destacamos os três corpora que integraram a pesquisa de Biber de análise multidimensional (BIBER, 1988, p.66): o Corpus Brown (1964), com um milhão de palavras do inglês americano escrito, o primeiro (BERBER SARDINHA, 2004a, p.7-8); o LOB (Lancaster-Oslo-Bergen, 1978), com um milhão de palavras do inglês britânico escrito; e o LLC (London-Lund Corpus, de 1980) com 500 mil palavras do inglês britânico falado).

Em segundo lugar, destacamos, por sua extensão atual, o Corpus of Contemporary American English (COCA), com mais de 400 milhões de palavras, desenvolvido por Mark Davis, da Brigham Young University. O COCA merece destaque por ser, de acordo com seu tutorial online, o maior corpus de inglês disponível gratuitamente. Os textos foram coletados e distribuídos igualmente entre falado, ficção, revistas populares, jornais e textos acadêmicos. Além disso, ele é semestralmente incrementado. Mark Davis disponibilizou esse corpus em uma plataforma online de forma que se possam fazer buscas por palavra, frases, lemas, partes do discurso (características morfossintáticas), colocados (palavras que aparecem próximas a uma palavra de busca) e

---

<sup>84</sup> informações retiradas do site <http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=products#world>

outras combinações, que podem ser observadas por gênero e ao longo do tempo. Embora substancialmente maior que o ANC, a desvantagem do COCA é que os textos não são disponibilizados em sua totalidade, apenas os resultados das buscas, restringindo seu uso, portanto, à interface disponibilizada por seus criadores.

Feitas essas considerações sobre corpus e LC, passaremos a apresentar teoricamente o modelo de análise multidimensional proposto por Biber (1988). Acreditamos que a observação de ocorrências de formas padronizadas interdependentes entre léxico e gramática, ou seja, entre as funções fonéticas, lexicais, morfológicas, sintáticas e semânticas, subjaz a proposta de uma metodologia de análise que leve em consideração todas essas características e como elas se relacionam entre si e em relação ao contexto de produção.

## 2.2. Análise Multidimensional

Nesta seção apresentaremos teoricamente a abordagem de análise multidimensional, no capítulo seguinte apresentaremos os procedimentos metodológicos propostos por essa abordagem, conforme foram seguidos nesta pesquisa.

A Análise Multidimensional, ou Análise Multitraço e Multidimensional de Variação de Registro<sup>85</sup>, ou simplesmente AMD, recebeu impulso principal (especialmente nos Estados Unidos) do linguista Douglas Biber, que também trabalha com análise de padronização (BERBER SARDINHA, 2004a). Refere-se a uma análise quantitativa a partir de levantamentos estatísticos de características linguísticas, realizadas por computador. Berber Sardinha apresenta-nos a seguinte definição:

*É uma abordagem para a análise de corpus que usa procedimentos estatísticos (principalmente análise fatorial), visando o mapeamento das associações entre um conjunto variado de características linguísticas dentro do corpus de estudo. Também usa procedimentos automáticos e semi-automáticos para análise do corpus, tais como etiquetagem morfossintática (part of speech tagging)*

(BERBER SARDINHA, 2004a, p.300)

---

<sup>85</sup> conforme proposto por Berber Sardinha (2004, p. 300), para *Multifeature Multidimensional Analysis of Register Variation*

Essa abordagem prevê a análise de variação textual a partir de diversas características linguísticas, por meio de “um procedimento estatístico usado para o agrupamento de dados com base em sua ocorrência” (p. 300), chamado Análise Fatorial.

Biber parte da perspectiva de que os falantes de uma língua possuem, além do conhecimento lexical, conhecimento sobre a estrutura e o uso da língua (BIBER, 1988, p. 8), para usá-la de acordo com as exigências funcionais e situacionais. Essa perspectiva também foi apontada em outros estudos como o de Francis (1995, In. HUNSTON; FRANCIS, 2000, p. 31), que lembra o papel inicial do léxico no processo comunicativo ao afirmar que não “escolhemos estruturas sintáticas e independentemente escolhemos itens lexicais para preenchê-las”. Na realidade, essas estruturas já fazem parte de conceitos comunicativos que possuímos, os quais “requerem itens lexicais centrais, e essas próprias escolhas encontram estruturas semânticas nas quais podem ser ditas com conforto gramatical”<sup>86</sup> (id. *ibid.*).

Por outro lado, a escolha se dá em consonância com a situação e com a função que o falante deseja desempenhar. Assim, a abordagem multidimensional considera que as características linguísticas representam diferentes aspectos de situações comunicativas. Esse caráter funcional foi apresentado também por Halliday (1991; HALLIDAY; HASAN, 1989) e adotado na pesquisa de Biber (1988), que distingue sete entre as principais funções de características linguísticas, cada uma das quais referem-se ao tipo e informação marcada no discurso: “ (1) ideacional, (2) textual, (3) pessoal, (4) interpessoal, (5) contextual, (6) processual e (7) estética” (BIBER, 1988, p. 34).

As duas primeiras funções são de cunho estritamente linguísticos. As funções ideacionais referem-se ao modo como as formas linguísticas representam o conteúdo proposicional (“tais como substantivos frequentes e frases preposicionadas ou um vocabulário altamente variado”<sup>87</sup>) (*ibid.*) e referencial. As funções textuais marcam estrutura e coesão, ou seja, características que refletem o tema, passivas, subordinadas, referências pronominais, demonstrativos, etc. As outras funções referem-se a como as

---

<sup>86</sup> Do original: “selecting syntactic structures and independently choosing lexical items to slot them. Instead, we have concepts to convey and communicative choices to make which require central lexical items, and these choices find themselves syntactic structures in which they can be said comfortably grammatically”

<sup>87</sup> Do original: “such as frequent nouns and prepositional phrases, or a highly varied vocabulary”

características linguísticas se relacionam com as informações exteriores ao texto. As funções pessoais denotam estilos individuais, atitudes em relação à mensagem, etc. Já as funções interpessoais denotam os papéis dos participantes, suas interações. As funções contextuais referem-se aos marcadores de tempo, de espaço, dos propósitos comunicativos e de percepção da cena. As funções processuais refletem os processos de produção e compreensão da mensagem. Finalmente, as funções estéticas refletem a preferência de escolha de formas, seja estabelecida por acordos linguísticos ou culturais.

As funções são observáveis, portanto, a partir da presença, recorrência ou ausência de determinados traços (ou características)<sup>88</sup> linguísticos, ou seja, pela variação<sup>89</sup> de traços linguísticos. Biber considerou, em sua análise, primeiramente as frequências de características linguísticas, ao contrário de investigações anteriores que partiam da observação situacional e funcional. O autor nos lembra, contudo que a observação da coocorrência de características linguísticas sozinha não é interessante, mas sim, a influência dos parâmetros situacionais ou funcionais pela sistematicidade de uso desses conjuntos de características linguísticas que coocorrem ao longo dos textos.

Estudos anteriores aos de Biber<sup>90</sup> objetivaram a identificação de características linguísticas específicas, marcando, geralmente, uma divisão unilateral de funções, isto é, ou um texto é representante de uma ou de outra. A proposta de Biber é inovadora pois defende que se deve levar em conta que textos, ao serem contrastados, podem ser mais ou menos, semelhantes em relação a diversas características e, portanto, devem ser analisados considerando-se diferentes dimensões.

Em outras palavras, é possível comparar dois textos por características como formal ou informal; interativo ou não interativo; ou ainda literal ou coloquial (BIBER, 1988, p. 9), que podem ser consideradas dimensões a partir da presença e recorrência de

---

<sup>88</sup> Conforme dissemos na introdução do trabalho, características ou traços linguísticos correspondem aos elementos lexicais e gramaticais típicos de um texto

<sup>89</sup> “Para Schiffrin (1994), a abordagem variacionista para o discurso baseia-se nos pressupostos de que ‘a variação linguística, isto é, a heterogeneidade, é determinada por padrões sociais e linguísticos’; ‘os estudiosos da variação tentam descobrir padrões na distribuição de maneiras alternativas de dizer a mesma coisa, ou seja, fatores sociais e linguísticos que são responsáveis pela variação nas formas de falar’ (Schiffrin, 1994: 282)” (PACHECO DE OLIVEIRA, 1997, p. 67)

<sup>90</sup> Biber cita os trabalhos Chaffé, 1982; “Olson, 1977; “Chafe and Danielewicz, 1986; “Stubbs, 1980, Brown and Yule, 1983; entre outros. (BIBER, 1988, p. 47)

determinados traços linguísticos. Todavia, a interpretação dessas características<sup>91</sup> nos leva a perceber que não existe uma oposição binária facilmente definida, especialmente quando maior número de textos é contrastado. Ou seja, podemos perceber que as qualidades não são exatamente ou uma ou outra, mas, um texto pode ser mais formal que outro, menos interativo que um terceiro texto e/ou menos literal, e a variação entre esses polos, sim, constitui cada dimensão.

Vejamos a definição de dimensão, segundo Berber Sardinha:

*Dimensão é o estatuto que um fator<sup>92</sup> assume assim que é interpretado do ponto de vista de sua função comunicativa. Uma dimensão permite visualizar características em comum partilhadas por uma porção significativa dos dados. A interpretação do fator leva em conta tanto as características linguísticas quanto as características partilhadas pelos registros que estão representados no fator. As dimensões permitem redefinir o quadro inicial de registros.*

(BERBER SARDINHA, 2004a, p. 304-305)

Para Biber (1988), um conjunto de características linguísticas que coocorre em diversos textos indica um uso padronizado e, conseqüentemente, definem uma dimensão linguística. O autor defende que “*cada dimensão compreende um grupo de características linguísticas coocorrentes e cada padrão de coocorrência pode ser interpretado em termos funcionais*”<sup>93</sup> (p. 14). A partir daí, avalia-se, empiricamente, a quantidade de dimensões, de funções relacionadas a cada dimensão e a importância de cada uma delas.

No caso das características linguísticas apontadas por Biber, podemos descrever um texto como, por exemplo, “*nominal versus verbal, ou estruturalmente complexo versus estruturalmente simples*”<sup>94</sup> (p.12). Biber apresenta três questões que nortearam sua investigação empírica:

<sup>91</sup> no modelo de Biber essas características são agrupadas estatisticamente em forma de fatores, especialmente porque se trata da análise de enorme quantidade de dados

<sup>92</sup> “conjunto de variáveis que coocorrem significativamente do ponto de vista estatístico” (BERBER SARDINHA, p. 304) [nota nossa]

<sup>93</sup> Do original: Each dimension comprises an independent group of co-occurring linguistic features, and each co-occurrence pattern can be interpreted in functional terms”

<sup>94</sup> Do original: “..., there are dimensions that compare texts in terms of their linguistic characterization, e.g., nominal versus verbal, or structurally complex versus structurally simple”

*(1) quais evidências podem sustentar essas impressões linguísticas? (2) Essas caracterizações representam uma única dimensão linguística, ou duas dimensões, ou mais de duas? Como um pesquisador consegue determinar quantas dimensões linguísticas são necessárias para dar conta da variação entre um conjunto de textos? (3) Há outras dimensões linguísticas que não foram representadas pelas impressões linguísticas acima? Se há, como elas podem ser descobertas?*<sup>95</sup>

(BIBER, 1988, p. 12)

Uma vez identificadas e interpretadas, as dimensões linguísticas “podem ser usadas para especificar as ‘relações textuais’ entre diferentes tipos de texto”<sup>96</sup> (id., p. 19). A ideia é a de quantificar com precisão a frequência de cada uma das características constituintes de cada dimensão presentes em cada texto, a fim de possibilitar a comparação entre textos em relação a cada dimensão simultaneamente. Segundo o autor, usar uma única dimensão pode provocar uma interpretação incompleta, “porque as relações entre textos não podem ser definidas unidimensionalmente”<sup>97</sup> (id., p. 20).

A metodologia proposta por Biber prevê portanto o caráter multidimensional de análise linguística, como vemos em suas próprias palavras; “a descrição da variação linguística de uma dada língua será multidimensional”<sup>98</sup> (BIBER, 1988, p. 22). É importante ressaltar que o objetivo do autor é descrever a língua como um todo, não apenas de um único gênero. Nesse sentido, a pesquisa que ora descrevemos não se assemelha à de Biber, por outro lado, nossos dados também englobam outros segmentos do inglês geral, ao utilizar pacotes lexicais (sequência de palavras recorrentes) extraídos do BNC, como será descrito no capítulo metodológico.

A seguir, descreveremos com mais detalhe a pesquisa de Biber (1988), a fim de que possam ser observados algumas decisões e procedimentos tomados pelo autor e que nortearam a nossa pesquisa.

---

<sup>95</sup> Do original: (1) What evidence can we give to support these linguistic impressions? (2) Do these characterizations represent a single linguistic dimension, or two dimensions, or more than two? How can a researcher determine how many linguistic dimensions are required to account for the variation among a set of texts? (3) Are there other linguistic dimensions that are not represented by the above linguistic impressions? If so, how can they be discovered?”

<sup>96</sup> Do original: “can be used to specify the “textual relations’ among different kinds of texts in English.” [destaque do original]

<sup>97</sup> Do original: “... because the relations among texts cannot be defines unidimensionally”

<sup>98</sup> Do original: “... the description of linguistic variation in a given language will be multi-dimensional”

Biber aplicou sua metodologia de análise a fim de descrever as relações textuais em amostras da língua falada e escrita, partindo do princípio de que existe pouca variação entre os dois registros. Os textos utilizados foram extraídos especialmente de corpora já estabelecidos; o LOB Corpus (Lancaster-Oslo-Bergen Corpus of British English) e o London-Lund Corpus of Spoken English, como mencionamos anteriormente. Biber acrescentou um conjunto de cartas pessoais e profissionais a fim de incluir material não publicado. Enfim, seu corpus de estudo foi composto por 17 variedades de textos escritos, entre os quais encontram-se amostras de reportagens (44 textos), editoriais (27 textos), textos religiosos (17 textos), prosa acadêmica (80 textos), ficção (74 textos subdivididos em geral, mistério, científica, aventura e romance); e 6 gêneros falados, tais como conversas face a face (44 textos), por telefone (27), discursos planejados (14) e espontâneos (16). No total, o corpus de estudo de Biber somou 481 textos com aproximadamente 960 mil palavras. Dada a grande quantidade de dados e a natureza da pesquisa, os textos foram submetidos a análises computacionais a fim de identificar a ocorrência de características linguísticas relevantes.

Primeiramente, considerou-se a inclusão de grandes quantidades de características linguísticas. Para que essas pudessem ser observadas o corpus foi etiquetado morfossintaticamente. Essa etiquetagem foi feita automaticamente e depois conferida manualmente. As categorias selecionadas foram divididas da seguinte forma: marcadores de tempo e aspecto verbal (passado, perfeito e presente); advérbios de lugar e tempo; pronomes (pessoais, impessoais, demonstrativos, indefinidos); interrogações (wh-questions); formas nominais (formas substantivadas, gerúndios, substantivos); passivas; estado (verbos ser, estar e haver); subordinação (complementos, participios, pronomes relativos, subordinadas adjetivas e adverbiais); preposições; adjetivos e advérbios; especificidades lexicais (razão forma/ocorrência, tamanho da palavra); classes lexicais (conjunções, salvaguardas, pronomes enfáticos, partículas do discurso, demonstrativos); modais (can, may, might, could, ought, should, must, will, would, shall); classes especiais de verbo (públicos, pessoais, persuasivos, perceptivos); formas reduzidas (contrações); coordenação e negação (BIBER, 1988, p. 211-245).



Em seguida, cada característica foi computada em termos de frequência total e por texto. Essas frequências foram normalizadas por 1.000 palavras, a fim de serem comparáveis entre os textos. A não normalização dos dados pode levar a observação inadequada, pois se um texto contiver menos palavras que outro, a quantidade de substantivos pode ser maior ou menor em relação a outros textos, por exemplo. Assim, a normalização de frequências apontaria, nesse caso, quantos substantivos existiriam se cada texto tivesse 1.000 palavras, deixando todos os textos no mesmo padrão de comparação. Pode-se assim observar a frequência média, mínima e máxima de cada característica (chamadas agora de variáveis) e como elas se distribuem entre os textos. A pesquisa de Biber retornou 67 contagens diferentes para cada um dos 481 textos. A partir dessa observação, Biber passou para a análise estatística chamada Análise Fatorial.

Segundo o autor, “a Análise Fatorial é a ferramenta estatística primordial da abordagem multi-característica/multidimensional de variação textual<sup>99</sup>”, pois ela reduz o enorme número de variáveis (as características morfossintáticas), agrupando-as conforme sua coocorrência (por meio da identificação das variáveis que se distribuem de forma semelhante), retornando um número de fatores a serem analisados.

A análise fatorial envolve a tomada de várias decisões e diversos procedimentos. Primeiramente é necessário decidir o método de extração dos fatores, pois a “análise fatorial em linguística é geralmente exploratória (mais que confirmatória)”<sup>100</sup> (BIBER, 1988, p. 82). Entre as diversas opções disponíveis Biber afirma que as mais comumente usadas são “principal factor analysis” (PFA), ou análise dos fatores principais e “principal components analysis” (PCA), ou análise dos componentes principais. Segundo o autor, a diferença entre as duas é que a segunda (PCA) “tenta considerar *toda* a variação nos dados” e a primeira (PFA) “tenta considerar apenas a variação

---

<sup>99</sup> Do original: Factor analysis is the primary statistical tool of the multi-feature/multi-dimensional approach to textual variation”

<sup>100</sup> Do original: “... the use of factor analysis in linguistics is usually exploratory (rather than confirmatory)”

*compartilhada*”<sup>101</sup>(Ibid.).

O autor escolheu o primeiro método de extração para a sua análise, o PFA, que retorna uma matriz fatorial. Depois disso, deve-se acrescentar um procedimento estatístico chamado de rotação oblíqua, que possibilita a interpretação dos dados, pois a matriz fatorial é de difícil interpretação. Biber optou pelo método de rotação Promax (BIBER, 1988, p. 85) que retorna valores positivos e negativos para as distribuições das características já agrupadas por fatores. Essa análise retornou sete fatores.

Os fatores também são apresentados na forma de um gráfico “*scree*”, a partir do qual os fatores são interpretados. A interpretação é feita assumindo que “a coocorrência reflete funções compartilhadas. Isto é, as características linguísticas coocorrem em textos, porque funcionam de maneira semelhante”<sup>102</sup> (BIBER et al., 1998, p. 147).

Primeiramente, o programa de análise fatorial, na fase de descrição estatística, retorna uma tabela com cada uma das características linguísticas selecionadas – no caso da pesquisa de Biber, 67 (e.g.: verbos no passado, verbos no tempo perfeito, verbos no presente, advérbios de lugar, advérbios de tempo, etc.) –, apresentando suas frequências (média, mínima, máxima, escala e desvio padrão) para cada um dos gêneros selecionados pelo pesquisador, sendo eles: reportagem, editorial, críticas de imprensa, religião, hobbies, sabedoria popular, biografias, documentos oficiais, prosa acadêmica, ficção em geral, ficção de mistério, ficção científica, aventura, romance, humor, cartas pessoais, cartas profissionais, conversas face a face, conversas por telefone, entrevistas, transmissões (*broadcasts*), falas espontâneas e discursos preparados.

Os textos foram, em seguida, confrontados a respeito da relevância de cada variável para cada um, confirmando a dimensão de análise. Para isso, são calculados os escores padronizados de cada texto. Isso é feito da seguinte maneira: Cada texto tem uma contagem de frequência de cada variável<sup>103</sup>. As variáveis presentes em cada dimensão

---

<sup>101</sup> Do original: The primary difference between principal factor analysis (PFA) and principal component analysis (PCA) is that PCA attempts to account for *all* of the variance in the data while a PFA attempts to account for only the *shared* variation.” [itálico do original]

<sup>102</sup> Do original: “... that co-occurrence reflects shared function. That is, linguistic features co-occur in texts because they function in similar ways”

<sup>103</sup> lembrando que variável refere-se a característica linguística

têm seus valores somados e esses valores são padronizados de acordo com os valores médios e de desvio-padrão. Assim, cada texto tem um valor fatorial, podendo ser negativo ou positivo, referente a cada dimensão. Felizmente, o analista não precisa fazer esses cálculos um a um, o próprio programa computacional de análise fatorial os faz automaticamente.

Assim, um novo conjunto de tabelas é extraído, apontando agora para os fatores, ou seja, agrupando as características recorrente e coocorrentes. No caso da pesquisa de Biber o programa retornou sete fatores, conforme reproduzidos a seguir e explicados em seguida:



FACTOR 4		FACTOR 5	
infinitives	.76	conjuncts	.48
prediction modals	.54	agentless passives	.43
suasive verbs	.49	past participial	
conditional		clauses	.42
subordination	.47	BY-passives	.41
necessity modals	.46	past participial	
split auxiliaries	.44	WHIZ deletions	.40
(possibility modals	.37)	other adverbial	
-----		subordinators	.39
-- no negative features --		(predicative adjs.)	.31
		-----	
		(type/token ratio	-.31)
FACTOR 6		FACTOR 7	
THAT clauses as		SEEM/APPEAR	.35
verb complements	.56	(downtoners	.33)
demonstratives	.55	(adverbs	.31)
THAT relative clauses		(concessive	
on object positions	.46	subordination	.30)
THAT clauses as		(attributive adjs.	.30)
adj. complements	.36	-----	
(final prepositions	.34)	-- no negative features --	
(existential THERE	.32)		
(demonstrative			
pronouns	.31)		
(WH relative clauses			
on object positions	.30)		
-----			
(phrasal coordination	-.32)		

Quadro 3: Resumo da estrutura fatorial obtida na pesquisa de Biber para os fatores 4, 5, 6 e 7 (BIBER, 1988, p. 103)

Os quadros 1 e 2 apresentam um valor para cada variável, chamado de carga (*loading*). A carga corresponde ao valor numérico que cada variável tem em relação à própria dimensão e às outras. Esse valor vai variar entre -1 e 1. Valores positivos indicam que a variável coocorre acima da frequência média e valores negativos indicam que a variável coocorre abaixo da frequência média. Quanto mais próximo de 0 menos significativa é para a dimensão. Portanto, Biber optou por não considerar valores inferiores a 0,30. Assim, podemos observar as variáveis mais significantes positivamente e negativamente. Julgamos importante apresentar uma breve interpretação dos fatores, a

fim de que se possam entender as decisões analíticas do autor, que culminaram na nomenclatura de dimensões por ele apresentada. No capítulo metodológico, apresentaremos uma descrição mais detalhada das características linguísticas presentes nos fatores obtidos em nosso estudo.

A interpretação de fatores exige que se considerem as funções compartilhadas pelas características coocorrentes em cada fator. O fator 1 (cf. BIBER, 1988, 104-108) é muito importante porque é caracterizado por 34 características linguísticas. Entre as características com valores negativos encontram-se substantivos (*nouns*), portadores de referenciais de significados e “alta frequência de substantivos indica grande densidade de informação”<sup>104</sup>(BIBER, 1988, 104); frases preposicionais, que geralmente também carregam informação; o tamanho das palavras e a razão entre ocorrências (palavras) e formas (*type/token ratio*), que denotam densidade lexical, ou seja, uso variado de vocabulário com significado específico; e adjetivos atributivos, usados para elaborar a informação nominal. Enfim, as cinco características estão associadas a comunicação com foco na informação.

As características com valores positivos no fator 1, segundo Biber (1988) podem ser associadas a um “foco envolvido e não interacional, devido a um propósito essencialmente interativo e afetivo e/ou [devido a] circunstâncias de produção altamente restritivas”<sup>105</sup> (BIBER, 1988, p. 105). Ou seja, entre as características com maiores valores estão: verbos no presente, que se referem ao contexto imediato ou à prosa informacional; verbos pessoais (ou privados), que expressam pensamentos, sentimentos; pronomes pessoais da primeira e da segunda pessoas, que denotam os participantes, normalmente em um discurso interativo; enfatizadores em geral, que demonstram a atitude do falante. As outras características marcam generalizações e fragmentações no texto, como as contrações, o “do”, como verbo auxiliar, que substitui uma frase ou oração em inglês; pronomes relativos, que substituem substantivos; preposições que indicam redução; pronomes demonstrativos e indefinidos que substituem referenciais

---

<sup>104</sup> Do original: “...and a high frequency of nouns thus indicates great density of information in a text.”

<sup>105</sup> Do original: “... an involved, non-informational focus, due to a primarily interactive or affective purpose and/or to highly constrained production circumstances”.

nominais; modais de possibilidade, que sinalizam para uma certa incerteza no discurso; negação analítica, verbo principal *be* e coordenação *não frasal* são usadas quando há baixa densidade de informação; a subordinação indica elaboração frasal, que pode ser associada a discursos específicos; *WH-clauses*, ou orações pronominais, “fornecem um jeito de falar sobre as perguntas”<sup>106</sup>; subordinação condicional e causativa podem ser associadas a afeto e exemplificação.

Enfim, o contraste entre as características com valores positivos e com valores negativos indicam dois parâmetros opostos, conforme apontado pelo autor:

*(1) o propósito principal do escritor/falante: informacional versus interativo, afetivo e envolvido; e (2) as circunstâncias de produção: essas circunstâncias caracterizadas por cuidadosas possibilidades de edição, possibilitando escolhas lexicais precisas e uma estrutura textual integrada, versus circunstâncias ditadas pelas restrições do tempo real, resultando em escolhas lexicais generalizadas e apresentação de informação geralmente de forma fragmentada*<sup>107</sup>

(BIBER, 1988, p. 107)

Após análise sobre esses dois parâmetros o autor chamou o fator 1 de “produção informacional versus envolvida”, conseqüentemente nomeando a dimensão 1.

A interpretação do fator 2 é menos extensa pois o fator é caracterizado por menos variáveis. As variáveis positivas “podem ser todas usadas para propósitos narrativos”. Isto é, narrativas no passado (verbos no passado) sobre um indivíduo não participante do enunciado (3<sup>a</sup>. pessoa), em forma de relato (verbos impessoais) que podem ocorrer de forma descritiva (gerúndio). As variáveis negativas, que se opõem às positivas, representam referência mais imediata pela presença de verbos no presente, já os adjetivos atributivos sugerem o uso elaborado de referenciais nominais em discurso não narrativo. Enfim, o fator apresenta uma distinção entre discurso narrativo e outros tipos de discurso; ao mesmo tempo em que apresenta oposição entre um discurso mais ativo, direcionado para o evento, e um discurso mais descritivo e expositivo. Assim, o autor

<sup>106</sup> Do original: “... provide a way to ‘talk about’ questions”.

<sup>107</sup> Do original: “(1) the primary purpose of the writer/speaker: informational versus interactive, affective and involved; and (2) the production circumstances: those circumstances characterized by careful editing possibilities, enabling precision in lexical choice and an integrated textual structure, versus circumstances dictated by real-time constraints, resulting in generalized lexical choice and a generally fragmented presentation of information”.

chamou a dimensão oriunda desse fator de “preocupações narrativas versus não narrativas” (BIBER, 1988, p.109)

O fator 3 é composto positivamente pelas formas relativas *WH*, que podem ser consideradas como referenciais, ou seja, elementos que identificam os referentes (substantivos) no texto. A coordenação e a nominalização coocorrentes indicam que o discurso referencial também “tende a ser integrado e informacional”<sup>108</sup> (BIBER, 1988, 110). No polo negativo encontram-se advérbios de tempo e de lugar e outros advérbios, que são referentes temporais, físicos e de modo. Considerando os polos negativo e positivo, o autor levanta a distinção entre referências explícitas, independentes ao contexto, e referências não explícitas, dependentes da situação. Assim, ele nomeou a dimensão de “referências explícitas versus dependentes da situação” (id., *ibid.*).

A interpretação das características presentes no fator 4, cujas variáveis todas carregam valores positivos, sugere a marcação de persuasão, pois a presença de modais de predição (*will*) indica certeza de acontecimentos futuros; a presença de modais de necessidade indica obrigação e necessidade; e a presença dos modais de possibilidade indica a possibilidade de algo ocorrer. Os “*suasive verbs*”, ou verbos do tipo “comandar, exigir, instruir” (BIBER, 1988, 111), indicam intenção sobre eventos no futuro. As subordinadas condicionais apontam as condições para que os eventos aconteçam. Os infinitivos, nos textos explorados pelo autor, indicam a atitude do falante, pois apareceram frequentemente acompanhados por adjetivos<sup>109</sup>. Auxiliares separados, verbos característicos deste fato, representam uma construção em inglês quando os advérbios são posicionados entre o verbo auxiliar e o principal, normalmente os auxiliares são modais, como os já mencionados. Enfim, o autor considera que a função compartilhada por essas características é “expressão explícita de persuasão”, como denominou a dimensão 4.

As características linguísticas presentes no fator 5 mostram a importância das formas passivas, as quais são usadas para dar maior ênfase ao assunto apresentado que

<sup>108</sup> Do original: “indicates that referentially explicit discourse also tends to be integrated and informational”

<sup>109</sup> “e.g., happy to do it; hped to see it” (id.*ibid.*)



ao agente. Essas construções são normalmente usadas em discursos abstratos, técnicos e mais formais. A análise também indica que subordinativas adverbiais e orações conjuntivas geralmente são usadas juntamente com passivas, implicando na complexidade dos discursos. O alto valor fatorial para a razão entre ocorrências e formas (*type/token ratio*), que denotam densidade lexical, indica que o discurso técnico é mais abstrato, marcado por uma densidade lexical relativamente baixa, frequentemente utilizando construções na forma passiva. O autor chamou a dimensão de “informação abstrata versus não abstrata” (cf. BIBER, 1988, 111-113).

O fator 6 é caracterizado no polo positivo especialmente por orações que são iniciadas por “*that*” em frases relativas, como complemento de adjetivos e de verbos, o que indica elaboração na informação, assim como preposições no final de orações e uso de demonstrativos. Por outro lado, sua coocorrência com preposições órfãs (*stranded prepositions*) sugere que são usadas para um discurso mais espontâneo, não planejado. O autor aponta que ele mesmo, em 1986, e Halliday, em 1979, já haviam observado esse tipo de característica no discurso oral, “porque a língua falada é produzida e compreendida como um processo em andamento, ela é caracterizada por um intrincado de movimentos [e por] complexas estruturas frasais com baixa densidade lexical”<sup>110</sup> (BIBER, 1988, p.113). Finalmente, o autor sugere que o fator 6 apresenta distinção entre um discurso que é informacional mas produzido em tempo real, denominando-o “elaboração informacional online”

Biber observa que qualquer análise para o fator 7 seria provisória, pois a maioria das variáveis presentes no fator também aparecem com valores maiores em outros. Assim o autor sugere que o “fator 7 não é forte o suficiente para uma interpretação firme” e opta por não utilizá-lo em seu estudo.

A análise fatorial levou Biber a seis dimensões de análise da língua inglesa, como vimos: (1) discurso com interação e envolvimento versus informacional e editado; (2) discurso narrativo, de relato versus não narrativo; (3) referências explícitas versus

---

<sup>110</sup> Do original: “... because spoken language is produced and comprehended as an on-going process, it is characterized by ‘an intricacy of movement [and by] complex sentence structures with low lexical density”

dependentes da situação; (4) expressão explícita de persuasão; (5) informação abstrata versus não abstrata; e (6) elaboração informacional online.

Como podemos ver pelos valores da análise fatorial, na maioria das dimensões dois conjuntos de características são predominantes. Isso é o mesmo que dizer que “*quando as características em um grupo ocorrem juntas frequentemente em um texto, as características no outro grupo são marcadamente menos frequentes naquele texto e vice versa.*”<sup>111</sup> (BIBER, et al., 1998, p. 147)

É importante destacar que num trabalho posterior Biber reduziu o número a cinco dimensões que considerou mais evidentes na caracterização multidimensional entre textos escritos e falados da língua inglesa. O autor manteve as quatro primeiras dimensões, excluiu a sexta e renomeou a quinta para “estilo impessoal versus não impessoal” (BIBER et al., 1998, p.183).

Biber contrastou também as dimensões em termos dos textos escritos e dos falados, a fim de verificar as semelhanças e as diferenças entre eles, seguindo as dimensões encontradas. Primeiramente, o autor observou as similaridades e diferenças por dimensão e, em seguida, considerou todas as dimensões simultaneamente (BIBER, 1988, p. 129-160).

A dimensão 1 apresenta essa diferença bem claramente, pois os textos com maiores escores em discurso com interação e envolvimento são os falados e os com maiores escores em produção informacional são os escritos. O que parece evidente, pois, normalmente, enunciados escritos são mais planejados que os orais.

A dimensão dois apresenta as amostras de língua escrita como mais representativas para os dois extremos de variação. Enquanto os textos de ficção são os de cunho mais narrativos, entre os menos narrativos encontram-se principalmente a prosa acadêmica e documentos não oficiais, assim como alguns registros escritos como os difundidos por mídia (broadcasts) e conversas por telefone.

---

<sup>111</sup> Do original: “That is, when the features in one group occur together frequently in a text, the features in the other group are markedly less frequent in that text, and vice versa.”

A dimensão três apresenta documentos oficiais e cartas profissionais com os maiores índices para referências explícitas, em oposição a broadcasts que apresentam os menores escores.

A quarta dimensão expressa marcas de persuasão que a princípio são comuns aos dois tipos de enunciados, falado ou escrito. Porém, os maiores escores ficaram com os registros do tipo cartas profissionais e editoriais, enquanto os menores escores ficaram com broadcasts e críticas de imprensa.

A quinta dimensão, da primeira análise de Biber, contrastou informação abstrata em oposição à não abstrata. Discursos mais abstratos são os mais técnicos que envolvem prosa acadêmica e documentos oficiais. Os menos abstratos são as conversas por telefone, ficção romântica e conversas face a face. Os outros registros apresentam valores intermediários, o que indica grande variação entre os registros nessa dimensão.

Finalmente, Biber conclui que as dimensões 1, 3 e 5 são as que apresentam maior distinção entre os gêneros escrito e falado, enquanto nas outras dimensões não há muita distinção entre ambos. Biber sugere que “cada dimensão está associada a um diferente conjunto de funções comunicativas e cada uma define um diferente conjunto de similaridades e diferenças entre os gêneros. É necessário considerar todas as dimensões para uma descrição adequada das relações entre textos falados e escritos”<sup>112</sup> (BIBER, 1988, p. 169). As dimensões apresentadas não representam todas as diferenças entre os gêneros estudados, mas oferecem uma base teórica para a análise contrastiva de gêneros em inglês, visto que sejam analisadas a partir das características linguísticas mais significativas compartilhadas pelas características situacionais e comunicativas.

A análise multidimensional também serviu de influência para diversos estudiosos da LC (KENNEDY, 1998; BERBER SARDINHA, 2000 e 2004a, 2004b ; CROSSLEY; LOWERSE, 2007) e foi replicado por Lee (2000) e pelo próprio Biber (BIBER; FINEGAN, 1989; BIBER, 1995). Entre os diversos estudos no Brasil que utilizaram a

---

<sup>112</sup> Do original: Each dimension is associated with a different set of underlying communicative functions, and each defines a different set of similarities and differences among genres. Consideration of all dimensions is required for an adequate description of the relations among spoken and written texts.

AMD (entre eles os de BERBER SARDINHA, 2004b<sup>113</sup>; e de SHERGUE, 2003<sup>114</sup>), citaremos os de Kauffmann (2005), que categorizou diferentes tipos de textos de jornal, conforme mencionamos anteriormente, e a pesquisa de Pacheco de Oliveira (1997) que investigou um corpus de composições em português e em inglês escritas por alunos, contrastando-as em relação a dimensões linguísticas.

Em sua pesquisa, Kauffmann (2005) coletou um corpus com 1.431 textos de variados gêneros jornalísticos, a saber: artigo, carta, chamada, coluna de notas, comentário, crítica, crônica, editoria, entrevista, nota de correção, notícia, reportagem e resenha. O autor ainda criou uma classificação que chamou de “outros”, a fim de poder incluir todos os textos das edições diárias do jornal do corpus de estudo “A Folha de S. Paulo”. As quase 500 mil palavras presentes no corpus de estudo foram etiquetadas automaticamente por suas características morfossintáticas e passaram pelos procedimentos de análise estatística fatorial propostos por Biber. Com o objetivo de verificar a variação linguística existente entre os gêneros do jornal, o pesquisador detectou duas dimensões de análise que colocam os textos entre os polos “Narrativo versus Expositivo” e “Argumentativo versus Informativo”, o que, conseqüentemente, reforçou o conceito de que “existe nos gêneros [do jornal] uma cisão básica entre opinião e informação”, além de evidenciar “a existência de uma instância narrativa também atuante” (KAUFFMANN, 2005, p. 121).

Pacheco de Oliveira (1997) contrastou três grupos de textos de alunos universitários escritos em português (L1)<sup>115</sup>, inglês (L1) e inglês como língua estrangeira (L2), em contextos culturais diferentes, Brasil e Estados Unidos, a fim de identificar e descrever dimensões textuais e caracterizar o gênero discursivo de composição de alunos universitários. Nesse aspecto essa pesquisa diferencia-se da de Kauffmann, que utilizou diversos gêneros, mas assemelha-se a nossa que também investiga características de um gênero específico. Assim, a autora selecionou 90 textos de cada grupo, totalizando 270

---

<sup>113</sup> Estudo que seguiu as dimensões propostas por Biber para investigar o gênero reunião de negócios

<sup>114</sup> estudo que usou a análise fatorial e a AMD para apontar as funções comunicativas e situá-las em textos da área médica

<sup>115</sup> L1 significa primeira língua, ou língua materna; enquanto L2 corresponde a segunda língua, ou língua estrangeira.

textos que foram submetidos a identificação e contagem manual dos traços linguísticos e, posteriormente submetidos a análise estatística fatorial. A autora também observou redes de formação textual (redes contextual, interacional, sentencial/intersentencial, organizacional, extensiva e texto) (PACHECO DE OLIVEIRA, 1997, p. 99). Após a extração e análise de fatores, estes indicaram para cinco diferentes dimensões de variação linguística em composições de alunos universitários: (1) estilo textual reduzido ou elaborado; (2) orientação discursiva interacional versus informacional; (3) explicitação do contexto situacional e cultural; (4) lógica coesiva e posicionamento pessoal; e (5) orientação acadêmica estrutural e informacional. Os dados também foram submetidos a outros testes, como a análise múltipla da variação (MANOVA) e análise de variação (ANOVA), para verificar a variação significativa entre os textos. Tais testes constataram que os textos estudados variam quanto à língua, quanto ao tópico e quanto à língua e ao tópico em interação. A pesquisadora concluiu que as diferenças linguísticas entre os textos podem ser atribuídas à língua (se L1 ou L2), à cultura ou “*background cultural*”; e, quando essas duas estão equilibradas, “as diferenças linguísticas são mais fortes” (PACHECO DE OLIVEIRA, 1997, p.223). Finalmente, baseando-se em seus achados, a autora propôs uma caracterização das composições quanto ao gênero discursivo, em termos funcionais, formais e culturais.

O estudo desenvolvido por Crossley e Lowrese (2007) chamou nossa atenção por utilizar a AMD na identificação de dimensões de análise linguística para padrões da língua, isto é, pacotes lexicais de duas palavras (bigramas). Os autores contrastaram a frequência de bigramas em nove corpora de inglês falado e dois de inglês escrito. A extração fatorial da frequência compartilhada entre os bigramas fez emergir quatro dimensões. A primeira, “*scripted versus unscripted*”, separa diálogos naturais de monólogos, textos escritos e diálogos oriundos de tarefas, contendo bigramas do tipo “*don’t know*” e “*you know*”. A segunda, “*deliberate vs. unplanned discourse*”, separa textos escritos e textos ‘memorizados’ de todos os outros tipos de texto falado presentes nos corpora estudados (e.g., “*you see*”, “*so I*” e “*the first*”). A terceira, “*spatial vs non-spatial discourse*”, foi marcada por elementos que marcam o espaço

geográfico separando-os dos outros tipos de textos falados e escritos (e.g., “*you go*” e “*have you*”). A quarta dimensão, “*directional vs. non-directional discourse*”, é marcada pelas características linguísticas que oferecem direção e administram o tempo (e.g., “*need to*”; “*back to*” e “*the other*”). Enfim, os autores sugerem um sistema para classificação de registro baseado em n-gramas lexicais e sintáticos, confirmando que os padrões estabelecidos por pacotes lexicais (bigramas) são fonte de informação sobre as estruturas linguísticas tanto de gêneros escritos como falados.

Apresentamos nesta seção a contribuição de Douglas Biber para a análise de características linguísticas entre textos escritos e falados em língua inglesa e ‘pincelamos’ sua repercussão em pesquisas no Brasil. A metodologia proposta pelo autor para Análise Multidimensional, como vimos, envolve procedimentos macros e micros. Os passos principais descritos nesta seção foram listados por Beber Sardinha (2004a, 305-306) e estão reproduzidos a seguir:

- Levantamento das características linguísticas relevantes por meio de ampla consulta à literatura disponível.
- Coleta ou adoção de um corpus de dados linguísticos representativo e compatível com as metas da análise.
- Transformação das características linguísticas em variáveis quantificáveis.
- Codificação dos dados baseada nas variáveis selecionadas, usando ferramentas computacionais para análise automática, semiautomática (interativa) ou manual.
- Conferência manual da codificação feita por computador para checar sua exatidão.
- Padronização das frequências (em geral por 1.000 palavras), para permitir a comparação entre variedades (textos, registros ou corpora) de extensões diferentes.
- Computação de frequências médias de cada variável.
- Análise Fatorial inicial, a fim de obter os pesos (*loadings*) de cada variável em cada variedade.
- Determinação do número de fatores por meio da aplicação de técnicas como observação dos valores *eigen (eigenvalues)* em um gráfico *scree (scree plot)*.
- Análise Fatorial posterior, com a rotação dos fatores.
- Cálculo de escores de cada texto por fator pela padronização dos escores com base na média e no desvio padrão.
- Cálculo de escores médios de cada variedade por fator.
- Interpretação de cada fator e rotulação das dimensões.

Na seção seguintes retomamos a discussão sobre gênero que, como vimos, é constante nas pesquisas de análise multidimensional como a nossa, apresentando sua relação com nosso corpus de estudo.

### **2.3. Gênero e letras de música**

Nesta seção discutiremos brevemente a questão do gênero por dois motivos principais. Primeiramente, por esta pesquisa se tratar de um estudo linguístico que se refere a estilos/gêneros musicais em forma de texto, o que requer que explicitemos nosso posicionamento linguístico. Em segundo lugar, como vimos nas seções anteriores, o conceito de linguagem subjacente à Linguística de Corpus, assim como as pesquisas que seguem o modelo de AMD, levam em consideração que as variações linguísticas ocorrem de acordo com o contexto, ou seja, estão atreladas ao gênero.

Mais uma vez é importante observar a sensibilidade da discussão sobre o conceito de gênero, sendo que permeia diversas áreas da Linguística de forma nem sempre consensual. Os critérios utilizados para a classificação genérica, normalmente abarcados pela tradição literária, podem ser vistos como conflitantes por serem de naturezas diversas, como por exemplo os que consideram aspectos relacionados à forma e ao conteúdo (o soneto), os aspectos relacionados à representação do real (o romance) ou os aspectos relacionados à organização do enunciado (o gênero carta) (KAUFFMANN, 2005, p.30). Outra problemática é que alguns trabalhos na área de LC oferecem também incoerências quanto ao uso dos termos gênero e registro (MAURANEN, 1998, apud. BERBER SARDINHA, 2003, p.2). O próprio Biber utiliza os termos indistintamente em seus primeiros trabalhos (1988), passa a utilizar a forma ‘registro’ nos trabalhos mais recentes (BIBER et al., 1998), até que, em seu livro de 2009 *Register, Genre, and Style* (BIBER; CONRAD, 2009), decide esclarecer melhor sua posição quanto a gênero e registro, privilegiando o termo ‘registro’.

Gênero, em nossa pesquisa, é visto como uma perspectiva de análise textual, na

qual são observados e descritos os propósitos comunicativos e o contexto situacional de textos, considerando as estruturas de construção textual como um todo. Adotamos essa perspectiva analítica da concepção de gênero, conforme proposta de Biber e Conrad (2009), em consonância com o modelo metodológico por nós escolhido.

A concepção de gênero proposta por Biber pressupõe o caráter probabilístico, semântico e funcional da língua, conforme discutido anteriormente neste capítulo, e sua relação com o contexto, apresentando semelhanças à concepção proposta pela linguística sistêmico-funcional (HALLIDAY; HASAN, 1989; HALLIDAY; 1991; SINCLAIR; 1991; HALLIDAY; WEBSTER; 2002). Segundo Halliday e Hasan (1989, p. 108) “um gênero é reconhecido pelos significados associados a ele; na realidade, o termo ‘gênero’ é uma forma abreviada para uma expressão mais elaborada, ‘potencial semântico específico de gênero’”<sup>116</sup>, sendo necessário que os textos preencham requisitos estruturais de cada gênero para serem considerados como integradores de determinado gênero. Essa exigência denota o caráter previsível do gênero, ou seja, é esperado de antemão que para pertencer a um gênero, um texto deva seguir uma ‘estrutura genérica potencial’. Halliday, portanto, não leva em consideração generalidades e padronizações de certas atividades sociais podem produzir.

Existe também uma perspectiva pedagógica para o conceito de gênero (SWALES, 1990) que enfatiza o propósito comunicativo dos textos. Mantendo o objetivo de ensino, gênero é visto “como um recurso para uma melhor compreensão dos aspectos intelectuais e esquemáticos” (SANTOS, 1996) presentes na realização de determinados discursos. Para Swales, (1990, p.58) os “gêneros envolvem uma classe de eventos comunicativos, de cujos os membros compartilham um conjunto de propósitos comunicativos (...) reconhecíveis pelos especialistas participantes da comunidade discursiva”<sup>117</sup>. O reconhecimento é possível porque os gêneros “compartilham padrões

---

<sup>116</sup> Do original: “a genre is known by the meanings associated with it; in fact the term ‘genre’ is a short form for the more elaborate phrase ‘genre-specific semantic potential’ ...”

<sup>117</sup> Do original: “genre comprises a class of communicative events, the members of which share some set of communicative purposes (...) are recognized by the expert members of the parent discourse community”.



de similaridade em termos de estrutura, estilo, conteúdo e público alvo”<sup>118</sup> (id.ibid.). Na visão do autor, o aprendiz de uma língua, seja materna ou estrangeira, torna-se mais competente quando se torna capaz de identificar as convenções e padrões linguísticos de diferentes gêneros.

O gênero está atrelado, portanto, às funções sociais e culturais desempenhadas pelo texto, de acordo com os propósitos comunicativos da comunidade discursiva. Nesse âmbito, o texto é considerado como a unidade básica de comunicação, ou seja, é a representação da “linguagem natural usada para comunicação, seja oral ou escrita”<sup>119</sup>, podendo “ser descrito de acordo com seu contexto” (BIBER; CONRAD, 2009, p.5) e seu caráter funcional (HALLIDAY; HASAN, 1989). O texto é, portanto, o objeto de análise do linguista, pois é nele que estão representadas as características linguísticas que denotam um registro.

Registro, por sua vez, “é o plano de expressão do gênero” (MARTIN, 1985, apud. BIBER; CONRAD, 2009, p.22), ou seja, apresenta a combinação de características linguísticas comuns a uma variedade de textos, de acordo com a situação de uso. Em outras palavras, para Halliday (HALLIDAY; HASAN, 1989, p.41) o registro é determinado pelo conjunto situacional/semântico e realizado pelas estruturas léxico-gramaticais identificadas num registro, a partir da busca por traços recorrentes presentes nos textos. (SANTOS, 1996). Pressupõe-se, portanto, que as características linguísticas centrais (substantivos, verbos) denotam a função representada a partir dos propósitos comunicativos e do contexto situacional do texto. Swales (1990) adverte que a investigação de traços linguísticos recorrentes não deve ser vista como “a busca por padrões normativos, mas como uma procura contínua das tendências operacionais de comunicação de um evento comunicativo pela comunidade comunicativa” (SANTOS, 1996).

Biber distingue as características de registro, gênero e estilo. Embora estilo não

---

<sup>118</sup> Do original “a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience”

<sup>119</sup> Do original: “... we use the term *text* to refer to natural language used for communication, whether it is realized in speech or writing.

faça parte do escopo desta pesquisa mantivemos a tabela conforme apresentada pelo autor:

<b>Característica definidoras</b>	<b>Registro</b>	<b>Gênero</b>	<b>Estilo</b>
Foco textual	amostras de textos	textos completos	amostras de textos
Características linguísticas	qualquer característica léxico-gramatical	expressões especializadas, organização retórica, formatação	qualquer característica léxico-gramatical
Distribuição de características linguísticas	frequentes e patentes nos textos da variedade	geralmente ocorrem um única vez, num lugar particular no texto	frequentes e patentes nos textos da variedade
Interpretação	as características servem a importantes funções comunicativas no registro	as características são convencionalmente associadas ao gênero: o formato esperado, mas geralmente não funcionais	as características não são diretamente funcionais; são escolhidos por seu valor estético

Quadro 4: Reprodução da tabela 1.1. Características definidoras de registro, gênero e estilo (BIBER; CONRAD, 2009, p. 16)<sup>120</sup>

As diferenças entre registro, gênero e estilo, segundo Biber e Conrad (2009), estabelecem-se levando em consideração o texto, enquanto unidade de análise; as características linguísticas analisadas; a distribuição dessas características; e sua interpretação.

No nível do gênero as características linguísticas são observadas em relação a seus aspectos estruturais no texto como um todo, focalizando-se, muitas vezes, mas não exclusivamente, em características linguísticas que ocorrem uma única vez, pois estas indicam como uma variedade de textos é construída.

Já no nível do registro e do estilo, as características linguísticas são observadas em relação a sua pertinência nos textos que representam uma variedade. Em suma, a análise de registro, que é o foco de Biber, “caracteriza as características linguísticas típicas de variedades textuais e conecta essas características funcionalmente ao contexto situacional

<sup>120</sup> Tradução nossa

da variedade”<sup>121</sup> (BIBER; CONRAD, 2009, p.16).

Associando essas perspectivas ao nosso corpus de estudo podemos dizer que cada uma das letras de música corresponde a um texto, pois cada letra de música corresponde a nossa unidade básica de análise, advinda de produção natural da língua. A análise desses textos vai apontar para características linguísticas típicas desse tipo de texto, que por sua vez, serão associadas a sua funcionalidade situacional (registro). Como nosso corpus é composto por mais de 6.000 textos, nossas descobertas devem apontar para características pertinentes a essa variedade textual como um todo (gênero).

Outro ponto importante é que a música popular desempenha uma função social e cultural, ou seja, possui propósito comunicativo de transmitir uma mensagem, esteja ela vinculada às exigências mercadológicas de consumo ou a situações de protesto ou política, ou outras. Essencialmente a música e sua letra têm a função de levar o ouvinte a ter experiências imaginárias, porém relacionadas ao mundo real, contribuindo para a construção do ser social.

É preciso lembrar que, embora nosso objetivo não seja definir um gênero “letras de música”, partimos do pressuposto de sua existência e esta pesquisa deve caracterizar linguisticamente a variedade de música popular selecionada. Embora alguns estudiosos considerem música como um tipo de poesia, nosso objetivo também não reside em observar estrutura métrica, rima ou criação poética. Acreditamos, como Costa (2002), que poesia e letras de música pertencem a gêneros diferentes que compartilham aspectos comuns em sua produção. Segundo o autor, “canção é uma peça verbo-melódica breve, de veiculação vocal” (COSTA, 2002, p.108), que embora seja caracterizada por elementos rítmicos, esses não são totalmente padronizados, pois, no caso do *rap*, por exemplo, “os padrões entoacionais são colocados de lado em prol da letra e do ritmo”

Esperamos encontrar semelhanças entre o discurso coloquial, a fala do dia a dia, no discurso da música popular para que, num outro momento (não neste trabalho),

---

<sup>121</sup> Do original: “... characterizes the typical linguistic features of text varieties, and connects those features functionally to the situational context of the variety.”

possamos utilizar essas características para o ensino de língua, pois como afirma Costa (2002, p.108), seguindo as ideias de Luiz Tatit (1996):

*Uma canção é uma fala camuflada em maior ou menor grau. Essa camuflagem consiste na transformação dos contornos entoacionais da fala pela estabilização do movimento frequencial de sua entoação dentro do percurso harmônico, pela regulação de sua pulsação e pela periodização de seus acentos rítmicos. Essa regulação confere, com efeito, uma padronização relativa da vocalidade, pois, se cada falante fala de um modo diferente, imprimindo uma gestualidade singular ao texto falado, a canção estabelece um determinado modo de vocalização desse texto.*

(COSTA, 2002, p.108)

Os elementos gestuais e entoativos da canção, ou das letras de música, confirmam seu caráter sociosemiótico, contudo, não serão abordados por nosso trabalho, uma vez que visamos o levantamento de características exclusivamente linguísticas pertinentes a esse gênero.

Neste capítulo apresentamos os princípios teóricos que ofereceram suporte para nossa pesquisa, mais especificamente, apresentamos a Linguística de Corpus e os pressupostos e conceitos que embasaram nosso trabalho. Em seguida, apresentamos teoricamente o modelo metodológico que selecionamos para a análise de nossos dados, a análise multidimensional, ou AMD. Uma questão que se estendeu ao longo de nosso texto foi o conceito de gênero. Assim, encerramos o capítulo tecendo uma breve explanação sobre seu conceito e como este se relaciona com esta pesquisa.

A seguir, a fim de esclarecer ao leitor sobre todos os passos de nosso trabalho, apresentamos nosso primeiro estudo, que chamamos de estudo piloto, realizado preliminarmente à nossa decisão de utilizarmos a metodologia de análise multidimensional. De fato, foi a partir desse estudo contrastivo entre nosso corpus de estudo e o inglês geral que percebemos a necessidade de uma pesquisa que nos fornecesse subsídios mais fundamentados sobre as características linguísticas de letras de música.

Após a apresentação do estudo piloto, retomaremos a pesquisa principal apresentando os passos metodológicos, os resultados obtidos e nossas considerações.

### 3. Estudo Piloto

No início de nosso trabalho, ao vislumbrarmos a possibilidade de trabalhar com Análise Multidimensional que usaríamos para descrever os fatores responsáveis pela variação linguística em letras de música popular anglo-americana, desenvolvemos um estudo piloto de convergência lexical entre letras de música compostas em inglês e inglês geral.

Partimos do princípio de que as letras de música, como dissemos na introdução, não podem ser consideradas simplesmente como poesias musicadas, pois embora poesia e letras de música compartilhem características semelhantes como métrica, rima, etc. (COSTA, 2002), as duas formas também apresentam características diferentes, especialmente pelo componente sonoro que acompanha as letras de música. Ademais, esperamos apresentar semelhanças das letras de música com a linguagem coloquial (BÉRTOLI-DUTRA, 2002; GRIFFITHS, 2003).

Sob tal perspectiva, desenvolvemos um estudo lexical contrastivo entre letras de música e a língua inglesa geral, cujos passos estão descritos a seguir.

A fim de atingirmos o objetivo principal de detectar similaridades e diferenças lexicais entre o discurso de letras de música popular anglo-americana e inglês geral, nosso estudo seguiu as seguintes orientações específicas:

- Descrição das frequências das palavras usadas no corpus de estudo.
- Descrição dos padrões léxico-gramaticais do corpus de estudo.
- Contraste dos padrões encontrados no corpus com os padrões existentes na língua inglesa, conforme representados no corpus de referência.

É importante ressaltar que este estudo piloto também encontrou sua fundamentação teórica principal na Linguística de Corpus. Como já apresentamos uma discussão sobre tais princípios teóricos no capítulo anterior, passaremos imediatamente a descrever os procedimentos metodológicos.

### 3.1. Metodologia

O objetivo do estudo piloto foi extrair, relacionar e contrastar padrões léxico-gramaticais (BÉRTOLI DUTRA, 2002; BERBER SARDINHA, 2004a; SINCLAIR, 1991; HUNSTON & FRANCIS, 1999) de um corpus composto por letras de música com corpora de inglês geral.

O objeto do estudo foi um corpus de letras de ‘música popular’ em língua inglesa, com pouco mais de 1 milhão de palavras, contrastado a dois corpora de referência da língua inglesa, que são o BNC (*British National Corpus*), com mais de 100 milhões de palavras em sua versão online, e o ANC (*American National Corpus*), com cerca de 15 milhões de palavras, em sua versão disponível online.

Cada um desses corpora está detalhado a seguir.

#### 3.1.1. Corpus de estudo

Nosso corpus de estudo, cuja composição está detalhada no quadro 6, é parte do corpus utilizado na pesquisa final, contendo, na fase piloto, 5.962 letras de música compostas originalmente em inglês, conforme expresso no quadro abaixo:

TOTAL	Ocorrências	1.078.882
	Formas	22.136

quadro 5: descrição resumida da composição do corpus de estudo

Como se pode observar, o corpus de estudo desta pesquisa possui pouco mais de um milhão de palavras, ou ocorrências (*tokens*), advindas de letras de música popular compostas originalmente em língua inglesa. Esses *tokens* são resultado do uso de 22.136 palavras diferentes, ou formas (*types*).

O corpus foi coletado levando em conta os seguintes critérios:

1. coletar a totalidade músicas compostas e gravadas por cada banda ou intérprete, desde que contivessem letras e que não se constituíssem de regravações também realizadas por outras bandas/artistas presentes no corpus.
2. selecionar artistas que fossem considerados representativos ou pela longevidade do conjunto de seu trabalho, ou por sua marca de sucesso em determinada época.

### 3. selecionar um subcorpus das músicas mais tocadas nos dias de hoje

Seguimos esses critérios porque acreditamos que, dessa forma, não estaríamos privilegiando músicas tocadas numa única época. Bandas como The Rolling Stones foram escolhidas por estarem presentes nas paradas de sucesso ao longo de seus mais de trinta anos de carreira(cf. *Billboard Charts*<sup>122</sup>). Por outro lado, bandas como a canadense Nickelback, por exemplo, foram selecionadas por pertencerem ao grupo das bandas que mais fazem sucesso nos anos 2000 (id.).

É importante ressaltar que algumas músicas ficaram propositadamente fora do corpus, como as essencialmente instrumentais (e.g. *A Saucerful of Secrets* e *Interstellar Overdrive*, da banda Pink Floyd). Algumas músicas dos Beatles, por exemplo, foram gravadas por vários intérpretes, e cantores como Frank Sinatra e Ray Charles gravaram, muitas vezes, a mesma música. Em casos como esses, a música foi mantida no corpus uma única vez.

As bandas e o número de canções gravadas por cada uma delas, que fizeram parte de nosso corpus de estudo, estão dispostas no quadro a seguir.

ARTISTA	QUANTIDADE DE CANÇÕES
Aerosmith	56
Beatles	250
Bon Jovi	195
Creed	69
Def Leppard	143
Elton John	326
Foo Fighters	109
Frank Sinatra	1.224
Greenday	169
Iron Maiden	215
The Killers	52
Led Zeppelin	79
Lenny Kravitz	125
Madonna	168
My Chemical Romance	61
Metallica	119

<sup>122</sup> Para ver detalhes sobre paradas de sucesso em diversas épocas e variadas categorias, sugerimos consultar os sites: <http://www.billboard.com/> e <http://www.musicimprint.com/Chart.aspx?id=C000002>

Nickelback	79
Nirvana	98
Paramore	50
Pearl Jam	146
Pink Floyd	164
Queen	172
Queensryche	112
Ray Charles	206
Red Hot Chilly Peppers	169
Rolling Stones	432
Simple Plan	49
Teen	202
TOP 100	121
U2	170

quadro 6: descrição detalhada do corpus de estudo

O corpus, portanto, foi composto por 5.962 músicas, compostas e/ou gravadas por 30 artistas diferentes. Optou-se por armazenar as letras de acordo com os compositores/intérpretes, levando em consideração a possível necessidade de se relacionar determinado padrão a um artista específico.

Procurou-se manter variedade de época de lançamento das músicas, além da inclusão de interpretes de cunho mais romântico como Frank Sinatra e Ray Charles, mais específicos dos anos de 1950 e 1960. Também foram coletadas as músicas mais tocadas pela MTV Brasil em agosto de 2008, assim como as músicas de artistas atuais voltados para o público adolescente, que chamamos de sub-orpus Teen, com artistas como Hannah Montanah, The Jonas Brothers e as trilhas sonoras dos filmes High School Musical I, II e III.

### 3.1.2. Corpora de referência

A pesquisa piloto também empregou, conforme já dissemos, dois corpora de referência: O BNC e o ANC.

O BNC (British National Corpus), como informamos, no capítulo teórico, é destaque na pesquisa com corpus por ter sido o primeiro a atingir a marca de 100 milhões de palavras e, especialmente, por ser o único disponível para compra em sua versão BNC World Edition.



O BNC World Edition, de acordo com seu tutorial online<sup>123</sup>, está dividido em 3.144 textos escritos e 910 textos falados, contém 4.054 textos e 100.467.090 palavras. Após a etiquetagem morfosintática (*POS tagging*) o número diminui levemente para pouco mais de 97.600.000 palavras (cf. <http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=numbers>).

A parte escrita do BNC inclui extratos de jornais regionais e nacionais, periódicos e jornais especializados voltados para o público de todas as idades e interesses, livros acadêmicos e ficção popular, cartas e memorandos publicados ou não, ensaios escolares e universitários, entre outros. A parte falada é composta por transcrições de conversas informais e coletadas de outras fontes como reuniões formais de negócios ou do governo, programas de rádio e conversas telefônicas. Podemos visualizar melhor essas informações no quadro a seguir, que apresenta a composição do corpus:

	textos	palavras
falado (demográfico) <sup>124</sup>	153	4.233.955
falado (contextualizado) <sup>125</sup>	755	6.175.896
escrito – livros e periódicos	2685	79.238.146
escrito para ser falado	35	1.278.618
escrito – miscelâneas	421	7.437.168

quadro 7: composição do BNC por categoria

O BNC também está disponível gratuitamente para buscas através de sites online. Embora não tenhamos especificamente usado tais ferramentas nesta pesquisa, consideramos mencionar que, se o pesquisador não tem a necessidade de observar os textos em sua totalidade, ele pode, por exemplo, consultar o site desenvolvido por Mark Davis da Brigham Young University (<http://corpus.byu.edu/bnc/>), cuja interface permite buscas por palavras ou frases no BNC, que retornarão a palavra em seu contexto de uso. A busca também pode ser refeita por classe de palavras ou padrões.

Trata-se da mesma plataforma utilizada para o COCA (Corpus of Contemporary American English), que descrevemos no capítulo teórico. Como dissemos, as buscas são eficientes e de extrema ajuda àqueles interessados no contexto de uso, porém, a interface não retorna listas de frequência de palavras ou bundles (agrupamentos de palavras) totais no corpus, que necessitamos para o desenvolvimento da pesquisa que ora descrevemos.

<sup>123</sup> Cf. <http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=numbers>

<sup>124</sup> Amostras de conversas naturais gravadas e transcritas.

<sup>125</sup> Palestras, comentários de notícias, reuniões de negócios, discursos políticos, sermões, comentários esportivos, etc.

O American National Corpus (ANC), ainda encontra-se em construção. Segundo seu tutorial online, é um projeto que pretende oferecer o quadro mais abrangente do inglês americano já criado, a fim de servir com fonte para a educação, para pesquisas linguísticas e lexicográficas, e para o desenvolvimento tecnológico. O objetivo é atingir o mínimo de 100 milhões de palavras de modo que o corpus possa ser comparável ao BNC.

Até o momento deste estudo piloto, esse corpus contava com 22 milhões de palavras também etiquetadas morfossintaticamente, das quais 15 milhões estão gratuitamente disponibilizadas para *download*<sup>126</sup>.

A vantagem do uso do ANC, para este estudo piloto, é que ele também disponibiliza gratuitamente, para *download*, listas de frequência por palavras individuais, bigramas (duas palavras que ocorrem juntas, e.g., “listas de”) e trigramas (três palavras que ocorrem juntas, e.g., “café com leite”). As listas apresentam, além das frequências, os lemas das palavras (sua forma base) e a etiquetagem morfossintática (*POS tagging*).

Descritos os três corpora que integraram nosso estudo, apresentaremos, a seguir, os procedimentos de nossa análise.

### 3.1.3. Procedimentos de Análise

A fim de desenvolvermos nosso estudo de convergência lexical, seguimos os seguintes passos para análise contrastiva:

1. Extração de listas de palavras de nosso corpus de estudo
2. Extração de listas de palavras do BNC
3. Obtenção de listas de palavras do ANC
4. Comparação manual das 500 palavras mais frequentes do corpus de estudo com os corpora de referência.

---

<sup>126</sup> Cf. <http://www.americannationalcorpus.org/>

Esse procedimento foi realizado levando em conta a frequência de palavras individuais e pacotes lexicais (*lexical bundles*) de três itens, ou trigramas.

Optamos pela utilização de trigramas por acreditarmos que representam a melhor combinação de palavras em uso. Nossa percepção corrobora a de Lafferty (LAFFERTY; SLEATOR; TEMPERLY, 1992) de que o “uso de uma palavra é determinado pela maneira que ela se associa à direita e à esquerda”<sup>127</sup> e a de Jelinek (1992) que justifica o uso de trigramas por primeiramente advirem de dados (“quanto mais melhor”) e por refletirem “simultaneamente a sintaxe, a semântica e a pragmática”<sup>128</sup> da palavra de busca.

Levando em consideração que nosso corpus de estudo possuía 22.136 palavras individuais, consideramos que o contraste das 500 palavras mais frequentes seria uma amostra representativa daquilo que é encontrado na música, em contraste ao inglês geral.

As 500 palavras mais frequentes de nosso corpus de estudo piloto, acompanhadas por suas frequências estão dispostas no quadro a seguir.

1 THE 43.428	126 THEN 1.419	251 MYSELF 598	376 DEAR 388
2 YOU 35.995	127 HOME 1.402	252 REMEMBER 596	377 ITS 387
3 I 35.964	128 ALWAYS 1.359	253 THOUGHT 596	378 UNTIL 386
4 TO 25.485	129 ABOUT 1.356	254 WAIT 589	379 LATE 382
5 AND 24.662	130 STILL 1.340	255 SOUL 587	380 FINE 381
6 A 23.138	131 HAD 1.333	256 LIVING 584	381 MORNING 379
7 ME 17.209	132 EVERY 1.327	257 COLD 581	382 YEARS 379
8 MY 14.635	133 KEEP 1.321	258 FEELING 580	383 ANYTHING 374
9 IN 13.966	134 WOULD 1.305	259 ONCE 575	384 TRYING 374
10 IT 13.118	135 OVER 1.287	260 SIDE 575	385 DOES 373
11 OF 12.685	136 US 1.261	261 PAIN 574	386 MISS 373
12 YOUR 10.694	137 EVER 1.255	262 THOSE 574	387 PRETTY 370
13 ON 9.951	138 HEAR 1.233	263 HANDS 566	388 TOUCH 370
14 THAT 9.471	139 AN 1.225	264 CHORUS 562	389 LAY 369
15 I'M 8.891	140 WERE 1.216	265 DONE 562	390 MONEY 369
16 ALL 8.729	141 YOU'LL 1.209	266 TILL 561	391 SUCH 366
17 BE 8.205	142 AM 1.185	267 HELP 559	392 NOBODY 365
18 IS 8.199	143 MUCH 1.146	268 LOOKING 559	393 MEET 364
19 FOR 7.759	144 MIND 1.145	269 DAYS 558	394 REASON 359
20 LOVE 7.606	145 BETTER 1.089	270 TOWN 553	395 RUNNING 358
21 DON'T 7.217	146 BELIEVE 1.087	271 WHAT'S 552	396 EASY 357
22 SO 6.492	147 OLD 1.085	272 BRING 550	397 GROUND 357
23 IT'S 6.220	148 NOTHING 1.083	273 GOODBYE 550	398 GUESS 355
24 BUT 6.157	149 INTO 1.074	274 NA 543	399 SOON 353
25 WE 5.865	150 THINGS 1.065	275 REAL 541	400 CHILD 351

<sup>127</sup> Do original: “A usage of a word is determined by the manner in which the word is linked to the right and to the left in a sentence”.

<sup>128</sup> Do Original: “Why should trigrams work so well? First, because they are firmly base don data, the more the better. Second, because they reflect simultaneously syntax, semantics, and pragmatics of the domain question.”

26 JUST 5.815	151 GONE 1.057	276 FOUND 538	401 BYE 348
27 KNOW 5.712	152 HOLD 1.037	277 FORGET 535	402 MATTER 347
28 NO 5.488	153 LIVE 1.018	278 THEY'RE 531	403 LISTEN 346
29 WITH 5.459	154 ALONE 1.011	279 LOSE 530	404 SOUND 346
30 WHEN 5.307	155 TRY 991	280 GOES 520	405 WENT 343
31 WHAT 4.956	156 I'D 988	281 KNEW 518	406 SAD 341
32 OH 4.905	157 GOTTA 964	282 ALRIGHT 517	407 UNDERSTAND 339
33 THIS 4.879	158 DREAM 963	283 MAKES 517	408 MUSIC 338
34 LIKE 4.842	159 THEM 963	284 BEST 515	409 NEXT 338
35 DO 4.550	160 LEAVE 946	285 START 512	410 RED 338
36 CAN 4.520	161 FACE 942	286 DEAD 507	411 SHAKE 338
37 GOT 4.447	162 RUN 940	287 WOMAN 505	412 READY 337
38 IF 4.386	163 YOU'VE 927	288 HEAVEN 504	413 PART 336
39 NOW 4.379	164 ANOTHER 925	289 LONELY 504	414 SEA 336
40 YOU'RE 4.253	165 LIGHT 915	290 ROCK 503	415 WORK 336
41 OUT 4.194	166 PLEASE 915	291 RAIN 500	416 FEAR 335
42 ONE 4.122	167 DID 911	292 DOOR 497	417 FUN 334
43 GET 4.024	168 SUN 907	293 CLOSE 496	418 BURNING 331
44 YEAH 4.024	169 HAS 906	294 WISH 496	419 TRUTH 331
45 DOWN 3.906	170 BEFORE 905	295 BECAUSE 495	420 ABOVE 330
46 GO 3.889	171 GOING 904	296 SMILE 492	421 FEET 330
47 UP 3.883	172 NEW 902	297 SEEN 490	422 GIRLS 330
48 TIME 3.820	173 SOMETHING 902	298 BLACK 489	423 LADY 330
49 ARE 3.613	174 REALLY 898	299 MAYBE 488	424 TOMORROW 330
50 NEVER 3.546	175 CALL 891	300 MEAN 485	425 WONDER 330
51 WAS 3.485	176 HEAD 884	301 KIND 484	426 BOYS 329
52 SEE 3.438	177 YES 882	302 TALK 484	427 EVERYONE 327
53 BABY 3.387	178 HAND 879	303 EVERYBODY 478	428 WANTS 327
54 SHE 3.387	179 INSIDE 876	304 CRAZY 477	429 TRIED 326
55 WILL 3.374	180 TONIGHT 871	305 HAPPY 475	430 WHITE 324
56 CAN'T 3.270	181 LAST 860	306 ROLL 475	431 MAMA 321
57 HAVE 3.260	182 PLACE 842	307 FLY 474	432 BROKEN 320
58 COME 3.249	183 STAY 840	308 ROUND 471	433 STRONG 318
59 NOT 3.194	184 OFF 838	309 OPEN 469	434 BABE 316
60 I'LL 3.181	185 THAN 837	310 TOLD 466	435 LIGHTS 316
61 WANT 3.103	186 THING 837	311 YOURSELF 465	436 FALLING 314
62 SAY 3.070	187 HIM 804	312 AH 462	437 SAYS 314
63 FROM 3.041	188 STOP 798	313 SOMEBODY 461	438 SET 312
64 TAKE 2.973	189 THEIR 791	314 SURE 461	439 WANTED 312
65 HER 2.933	190 LET'S 788	315 MANY 460	440 DARK 311
66 WAY 2.851	191 DIE 784	316 USED 460	441 C'MON 306
67 THEY 2.822	192 WALK 780	317 CHANCE 459	442 BORN 305
68 BACK 2.734	193 WE'LL 780	318 CAME 457	443 COULDN'T 305
69 GONNA 2.733	194 MINE 778	319 ENOUGH 457	444 ASK 304
70 AS 2.731	195 FALL 777	320 LINE 457	445 MOTHER 302
71 MAKE 2.729	196 CRY 774	321 FRIEND 456	446 WHOA 302
72 AWAY 2.689	197 MADE 774	322 TEARS 454	447 LAND 299
73 THERE 2.661	198 OOH 774	323 MOON 453	448 WHO'S 298
74 AT 2.642	199 TWO 770	324 SOMETIMES 452	449 WILD 298
75 LET 2.575	200 TRUE 768	325 FRIENDS 451	450 LORD 297
76 I'VE 2.425	201 TURN 768	326 FOREVER 450	451 SEEM 297
77 DAY 2.382	202 SHOULD 767	327 FIGHT 448	452 GAME 296
78 HOW 2.320	203 SOMEONE 766	328 ARMS 444	453 KILL 292
79 NIGHT 2.306	204 WRONG 760	329 RIDE 444	454 LOW 291
80 HEART 2.260	205 SWEET 758	330 WORDS 444	455 BEAUTIFUL 290
81 HE 2.223	206 LA 757	331 LIE 443	456 STANDING 290
82 LIFE 2.221	207 HARD 756	332 GETTING 441	457 HATE 288
83 FEEL 2.178	208 WITHOUT 756	333 TIMES 440	458 WHOLE 288
84 TELL 2.079	209 HE'S 751	334 STAR 439	459 CITY 287
85 RIGHT 2.061	210 MUST 737	335 KNOWS 438	460 TAKES 286
86 COULD 2.055	211 DANCE 736	336 TODAY 438	461 SHINE 284
87 WELL 2.051	212 LOST 733	337 YOUNG 438	462 WAR 283
88 HERE 2.046	213 EVERYTHING 729	338 ANY 437	463 HEARTS 282
89 CAUSE 2.041	214 LEFT 726	339 HOPE 437	464 HOT 282

90 THERE'S 1.938	215 COMING 725	340 ALIVE 435	465 HIT 281
91 TOO 1.937	216 FREE 725	341 HIDE 434	466 LIPS 279
92 MORE 1.932	217 SONG 720	342 HELL 432	467 WALKING 279
93 WHERE 1.871	218 COMES 718	343 BLOOD 430	468 DEATH 278
94 BY 1.845	219 END 710	344 STARS 429	469 MEN 277
95 MAN 1.836	220 SING 710	345 EACH 428	470 PARTY 277
96 LITTLE 1.831	221 PEOPLE 709	346 SLEEP 426	471 GUN 276
97 GOOD 1.823	222 SHOW 703	347 MIGHT 424	472 HA 275
98 NEED 1.820	223 CARE 699	348 AFTER 422	473 DOESN'T 274
99 BEEN 1.814	224 WHILE 698	349 DIDN'T 420	474 GETS 274
100 HEY 1.775	225 THESE 695	350 LIES 420	475 SINCE 273
101 OR 1.727	226 BAD 689	351 MOVE 420	476 BURN 272
102 GIVE 1.694	227 HIGH 683	352 ALONG 418	477 GAVE 272
103 THAT'S 1.663	228 SAME 679	353 WIND 414	478 HURT 272
104 THINK 1.655	229 WAITING 676	354 SEEMS 411	479 AIR 270
105 WORLD 1.652	230 OWN 670	355 BEAT 409	480 LIVES 267
106 OUR 1.643	231 PLAY 668	356 TOOK 409	481 SOMEWHERE 267
107 EYES 1.639	232 SKY 659	357 SAVE 408	482 TIRED 267
108 AGAIN 1.632	233 BREAK 658	358 HONEY 407	483 LEARN 266
109 ONLY 1.628	234 KISS 658	359 STREET 402	484 SOMEDAY 266
110 THROUGH 1.614	235 PUT 657	360 THOUGH 402	485 BED 263
111 WHO 1.604	236 NAME 649	361 YOU'D 402	486 BIT 259
112 WANNA 1.553	237 CHANGE 647	362 WATCH 401	487 PAST 259
113 SAID 1.547	238 DREAMS 645	363 UNDER 400	488 UPON 258
114 LONG 1.545	239 BOY 639	364 FOOL 397	489 TIGHT 256
115 WHY 1.542	240 BLUE 633	365 VERY 397	490 DARLING 255
116 GIRL 1.534	241 GOD 619	366 DEEP 396	491 FULL 255
117 WE'RE 1.533	242 MAY 618	367 BEHIND 395	492 BLUES 254
118 AROUND 1.517	243 STAND 618	368 ROOM 395	493 CONTROL 252
119 SOME 1.517	244 BIG 614	369 WORD 395	494 LOVED 251
120 HIS 1.496	245 TOGETHER 606	370 SAW 394	495 USE 251
121 SHE'S 1.494	246 FAR 604	371 WE'VE 391	496 STEP 250
122 AIN'T 1.478	247 OTHER 604	372 ELSE 390	497 MOMENT 249
123 FIND 1.471	248 EVEN 602	373 FIRST 390	498 EYE 248
124 WON'T 1.471	249 YA 602	374 HEARD 389	499 HELLO 248
125 LOOK 1.437	250 FIRE 600	375 ROAD 389	500 LOVING 248

quadro 8: 500 palavras mais frequentes no corpus do estudo piloto

Todas as palavras acima, logicamente, aparecem no BNC e no ANC. Então, a fim de verificarmos a frequência, em relação aos corpora de referência, passamos a investigar as palavras também nesses corpora, contrastando suas frequências.

Para que os corpora se tornassem comparáveis, normalizamos as frequências dos três por 100. Assim, por exemplo, o item mais frequente dos três corpora, “*the*”, que aparece em nosso corpus de estudo (CE) 43.428 vezes, terá a frequência normalizada para 4,025278019, simplesmente, 4,025. No BNC, o item “*the*” aparece 6.055.105, com a frequência normalizada, fica com o valor 6,026953702, ou, 6,027. E, no ANC, esse item aparece 1.204.816, cuja frequência normalizada exprime o valor 5,435672526 ou 5,44. Em outras palavras, se cada um dos corpora utilizados tivesse apenas 100 palavras, o “*the*” teria aparecido cerca de 4 vezes no CE, 6 vezes no BNC e 5 vezes no ANC. Em suma, esse procedimento faz com que as frequências compartilhem o mesmo

parâmetro de comparação. Para se normalizar as frequências faz-se o seguinte cálculo para cada palavra:

$$(\text{número de frequência} \div \text{número de palavras do corpus}) \times 100$$

Esse cálculo foi feito por meio de planilhas no programa Excel, pois os dados brutos são armazenados e transformados facilmente. Em seguida, contrastamos as frequências nos três corpora.

A fim de não nos concentrarmos unicamente em palavras isoladas, realizamos o mesmo procedimento com trigramas. Ou seja, extraímos listas de frequências de três palavras ocorrendo juntas (e.g. “*I want you*”; “*I love you*”; “*I don’t know*”), selecionando os 500 trigramas mais frequentes do CE, no BNC e no ANC, normalizamos e contrastamos suas ocorrências. Os 30 trigramas mais frequentes nos três corpora estão dispostos no quadro 11, na seção seguinte.

Essas frequências foram contrastadas e analisadas e os resultados de nossa análise estão dispostos a seguir.

### 3.2. Resultados

Após coletadas e normalizadas as frequências das palavras e coletados os trigramas dos três corpora, os dados foram contrastados a partir do CE. Ou seja, manualmente, procuramos nos corpora de referência a existência, ou coocorrência, das palavras, contrastando suas frequências, assim como a ocorrência dos trigramas nos três corpora.

Por se tratar de uma enorme quantidade de dados, trabalhamos com as 500 palavras e os 500 trigramas mais frequentes, mantendo nossa perspectiva de identificar traços de similaridade.

Verificamos que todas as 500 palavras do CE aparecem nos corpora de referência também entre as principais palavras. As 6 primeiras palavras mais frequentes são exatamente as mesmas no ANC e no BNC. Porém, em nosso CE a frequência das palavras apresenta leves alterações.

A palavra mais frequente nos três corpora é “*the*”, com frequências normalizadas

para 4,025 no CE; 6,027 no BNC e 5,44 no ANC.

Já a nossa segunda palavra mais frequente, “*you*” (cf. quadro 9), no CE aparece 3,336, em oposição a 0,585 no BNC e 0,80 no ANC, considerados em sua totalidade, ou seja, parte oral e parte escrita. Vejamos as 15 primeiras no quadro a seguir:

palavra	frequência CE	frequência BNC	frequência ANC
1. THE	4,02	6,02	5,44
2. YOU	3,33	0,58	0,80
3. I	3,33	0,73	0,85
4. TO	2,36	2,58	2,40
5. AND	2,28	2,61	2,68
6. A	2,14	2,17	2,21
7. ME	1,59	0,13	0,15
8. MY	1,35	0,14	0,24
9. IN	1,29	1,93	1,84
10. IT	1,21	0,91	1,15
11. OF	1,17	3,03	2,73
12. YOUR	0,99	0,13	0,11
13. ON	0,91	0,72	0,63
14. THAT	0,87	1,04	0,76
15. ALL	0,80	0,27	0,23

quadro 9: contraste de frequência das palavras mais frequentes no CE com o BNC e o ANC

Como podemos observar, as palavras ocorrem nos três corpora, porém com frequências variáveis. Fica bem evidente que no caso das letras de música presentes no CE, a presença dos pronomes pessoais de primeira e segunda pessoas (*I*; *you*) marcam a preferência por um discurso interpessoal entre a pessoa que canta (eu) e o ouvinte (você). Isso já havíamos verificado em nossa pesquisa de mestrado (BÉRTOLI-DUTRA, 2002).

Quando observamos palavras que exercem função básica, como substantivos, verbos, pronomes, etc., podemos observar variações entre os corpora, como as demonstradas na amostra a seguir:

palavra	frequência CE	frequência BNC	frequência ANC
'm	0,82	0,062	0,15

be	0,76	0,64	0,48
is	0,75	0,96	1,02
love	0,70	0,02	0,01
we	0,54	0,29	0,42
know	0,53	0,11	0,21
like	0,44	0,14	0,02
do	0,42	0,17	0,09
can	0,41	0,21	0,21
got	0,41	0,08	0,06
if	0,40	0,25	0,25
one	0,38	0,28	0,0018
get	0,37	0,95	0,10
down	0,36	0,91	0,05
go	0,36	0,86	0,021
up	0,36	0,20	0,017
time	0,35	0,15	0,010
are	0,33	0,45	0,47
never	0,32	0,05	0,059
was	0,32	0,85	0,71
see	0,31	0,11	0,104
baby	0,31	0,0079	0,008

quadro 10: amostra de contraste de frequência de itens lexicais do CE com o BNC e o ANC

Ao contrastarmos quantitativamente a ocorrência de palavras nos corpora, podemos identificar que a presença das formas “*m*” e “*are*” acompanha o uso pronominal destacado anteriormente. Especialmente a forma de primeira pessoa, que apresenta frequência muito maior em nosso corpus. A palavra “*are*” pode acompanhar outras formas, não estando seu uso atrelado a um pronome, diferente de “*m*”, que só aparece quando acompanhado de “*I*” e, por isso, sua frequência aparece de forma mais equilibrada, quando contrastada aos outros corpora.

O uso não flexionado do verbo “*be*” também aparece de forma equilibrada, assim como “*is*”, “*are*” e “*we*”.

Palavras como “*love*”, “*no*”, “*like*”, “*do*”, “*can*”, “*got*”, “*if*”, “*one*”, “*up*”, “*time*”, “*never*”, “*see*” e “*baby*” são sobreusadas no corpus de estudo, especialmente



“*baby*” e “*one*”. Vale ressaltar que a palavra “*one*” é bem menos usada no corpus americano que no britânico.

Esse comportamento acontece em sua quase totalidade para as 500 palavras mais frequentes no CE, ou seja, nossa análise de palavras individuais indica que, em termos quantitativos, a escolha de itens lexicais em letras de música se assemelha à escolha de realizada em outras formas da língua inglesa, o que faz com que nos arrisquemos a dizer que linguagem usada em letras de música pode ser considerada como coloquial, pois está presente em outros gêneros da língua inglesa. Vale ressaltar que os corpora de referência não possuem letras de música entre seus dados.

Acreditando que apenas essa análise não seria o suficiente, passamos a procurar a frequência de trigramas bem formados em língua inglesa, presentes no CE e também nos corpora de inglês geral.

A extração de trigramas do CE nos rendeu uma ocorrência de 129.117 dos mesmos. Já o BNC apresentou 5.431.734 trigramas diferentes. Esses trigramas foram extraídos a partir dos dados advindos dos textos, por meio de uma sequência de programação por nós desenvolvida. Os trigramas do ANC foram baixados do próprio site do ANC, subdivididos em trigramas do ANC falado (1.453.050) e trigramas do ANC escrito (4.236.030).

A seguir apresentamos uma tabela com os 30 trigramas mais frequentes nos três corpora envolvidos em nosso estudo, apresentando a subdivisão do ANC:

trigramas CE	trigramas BNC	trigramas ANC falado	trigramas ANC escrito
1. I WANT TO 540	ONE OF THE 20760	I I DO 997	THEUNITEDSTATES2600
2. I LOVE YOU 505	THE END OF 11604	AND I AM 993	THE NEW YORK 2477
3. I DON'T KNOW 371	AS WELL AS 11502	YOU THINK OF 99	NEW YORK TIMES 2427
4. LA LA LA 333	PART OF THE 10182	YEAH YEAH OH 99	ONE OF THE 2217
5. I DON'T WANT 294	THERE IS A 9655	YEAH BUT UM 99	THE THE THE 2141
6. DON'T WANT TO 285	OUT OF THE 9162	WAS NICE TALKING 99	THE OF THE 1635
7. OH OH OH 256	SOME OF THE 9142	UM THERE IS 99	IT IS A 1554
8. YEAH YEAH YEAH 254	EQUO HE SAID 8922	TIME AND I 99	THE IN THE 1505
9. I KNOW THAT 239	A NUMBER OF 8358	THAT IS SO 99	I DO NOT 1313
10. I DON'T WANNA 228	END OF THE 7931	TAKING CARE OF 99	IT IS NOT 1213
11. I GOT A 214	THERE WAS A 7741	SIT DOWN AND 99	THE THE A 1156
12. IN LOVE WITH 207	IT WAS A 7685	RIGHT BUT UH 99	IN THE THE 1124
13. I WANT YOU 205	THE FACT THAT 7673	NOT KNOW THE 99	THE A THE 1112
14. NA NA NA 202	THERE IS NO 7457	IT THAT WAY 99	A THE THE 1061
15. WANT TO BE 200	BE ABLE TO 7298	IT FOR A 99	A IN THE 1056
16. YOU AND ME 195	IN ORDER TO 7110	IT A LITTLE 99	OF THE NEW 1049
17. YOU WANT TO 191	TO BE A 6992	IS I AM 99	THERE IS A 1006
18. YOU AND I 187	IT IS NOT 6893	IUNDERSTAND THAT 99	AS WELL AS 980

19. IN YOUR EYES 184	PER CENT OF 6841	I THINK ABOUT 99	A LOT OF 958
20. I WANNA BE 183	A LOT OF 6574	I REALLY I 99	THE WHITE HOUSE 948
21. I DON'T NEED 181	IT IS A 6377	I LISTEN TO 99	THE END OF 934
22. IF YOU WANT 180	EQUO BQUO I 6291	I I UH 99	A A A 879
23. AND I DON'T 178	HE SAID BQUO 5929	I HAD THE 99	THERE IS NO 871
24. IN MY HEART 177	IN TERMS OF 5816	I GUESS THERE 99	THE THE IN 865
25. AND I KNOW 176	IT WOULD BE 5764	I AM LOOKING 99	THE THE OF 818
26. THIS IS THE 175	AT THE END 5608	HUM UM I 99	OF THE THE 815
27. CAN'T YOU SEE 174	MOST OF THE 5403	HUH AND UM 99	THE A A 812
28. TO BE A 169	ON THE OTHER 5391	GOOD THAT IS 99	SOME OF THE 795
29. YOU KNOW THAT 167	THAT IT IS 5389	AND UH MY 99	PART OF THE 782
30. ON AND ON 166	THE NUMBER OF 5372	AND THERE ARE 99	THE TO THE 775

quadro 11: 30 primeiros trigramas no CE contrastados com trigramas dos corpora de referência

O quadro 11 apresenta 30 trigramas presentes nos corpora de referência e no corpus de estudo, a fim de que o leitor possa visualizar o tipo de trabalho que realizamos. A partir da lista dos 500 trigramas mais frequentes do CE, manualmente observamos sua presença nos outros corpora. Essa observação não levou, a princípio, em conta a frequência dos trigramas nos outros corpora, apenas sua ocorrência.

Dos 500 trigramas mais frequentes no CE, apenas 12 não estão presentes em um ou outro dos corpora de referência (consideramos as versões falada e escrita do ANC juntas, para esta análise). Os 222 primeiros trigramas estão presentes tanto no corpus de estudo como em ambos os corpora de referência, BNC e ANC. A partir do 223º trigramas (*ooh ooh ooh*), os outros onze começam a aparecer exclusivamente no CE. São eles: “*c'mon c'mon c'mon*”; “*ooh ooh ooh*”; “*oo oo oo*”; “*good to ya*”; “*aah aah aah*”; “*wanna be your*”; “*need your love*”; “*love love love*”; “*all I wanna*”; “*ah ah ah*”; “*be your man*”.

Podemos observar que sete dessas expressões apresentam variações da escrita que não condizem com a forma culta, como a repetição de letras (e.g. “*aah*”; “*ooh*”) que é usada, muitas vezes em letras de música, para indicar o som prolongado ao cantar; as formas contraídas como “*wanna*” (equivalente a “*want to*”) e “*ya*” (equivalente a *you*); e a repetição de palavras (*c'mon c'mon*).

O caso específico de “*need your love*” não aparece nos corpora de referência, para nossa surpresa, apenas expressões como “*need your advice/passport/help/expertise/attention*”, o que pode justificar o sobreuso da palavra “*love*” apontado na análise de palavras individuais. O mesmo ocorre com “*be your man*”, que não foi encontrado nos corpora de referência, apenas expressões como “*be*

*your dream/friend/guide/home/idea/you*”. Podemos inferir que essa expressão parece ser de uso mais comum em letras de música por também apresentar referência a amor e sexualidade.

Por outro lado, o trigrama mais frequente do CE, “*I want to*”, ocorre 540 vezes, o que equivale a 0,5% das vezes e no BNC 2.110 vezes, equivalendo a 0,02%. Embora pareça uma grande diferença, a própria presença do trigrama já confirma seu uso frequente na língua inglesa em geral, não apenas nas músicas. Interessantemente, expressões como “*la la la*” aparecem nos três corpora. Isto é, embora possam parecer de uso exclusivo da música, também são usadas no inglês geral. Dados esses resultados, apresentamos a seguir nossas conclusões sobre o estudo piloto.

### 3.3. Conclusão

A constatação de que as palavras mais frequentes em nosso corpus de estudo também se encontravam, de uma maneira geral, entre as mais frequentes do inglês geral, seja ele americano ou britânico, confirma o que já havia sido identificado em nossa pesquisa de mestrado. (BÉRTOLI-DUTRA, 2002). Por outro lado, o estudo indicou a mesma tendência quando consideramos três palavras juntas ou trigramas, o que não tínhamos verificado em nosso estudo anterior.

O sobreuso de palavras como “*baby*” e “*love*” indica a tendência ao romantismo presente nas músicas, sugere um tipo de “fala à pessoa amada”. O que também pode ser confirmado pelo uso exclusivo dos trigramas “*be your man*” e “*need your love*”. Vale ressaltar a possibilidade, que não verificamos, de que a maior parte das ocorrências de “*baby*” nos corpora de referência (ANC e BNC) seja em relação a um bebê, e não como uma forma de tratamento afetuoso, que é a mais frequentemente usada no CE.

Este estudo nos levou a concluir que somente uma análise quantitativa de frequências normalizadas não seria o suficiente para apontar as características das letras de música. Um dos fatores que justificam essa conclusão é a não etiquetagem do corpus de estudo. Só pelo índice de frequência, por exemplo, não podemos saber quando

palavras como “*love*”, “*like*” ou “*know*”, bastante frequentes no CE, estão sendo usadas como verbo ou substantivo, o que, acreditamos, poderia nos dar uma ideia melhor de como as letras se comportam. Isto é, se as letras usam a forma “*know*”, por exemplo, para indicar o que se sabe ou conhece da mesma forma que no inglês geral ou se está sendo usada mais frequentemente para preencher espaços conversacionais, com expressões como “*you know*”.

Observar a ocorrência de trigramas mostrou-se bastante interessante, pois conseguimos apontar uma nova perspectiva de aproximação do discurso geral com o das letras de música e, portanto, poderíamos indicar satisfatoriamente o uso da música para o ensino de línguas – que era uma possibilidade que vislumbrávamos durante toda a execução da análise.

Ao mesmo tempo em que as duas esferas de análise mostram-se bastante interessantes e reveladoras, elas nos provocaram a necessidade de descrever o discurso da música mais detalhadamente e, para isso, seria necessário a etiquetagem do corpus e um outro tipo de abordagem metodológica.

Enfim, o estudo piloto impulsionou o estudo principal cuja metodologia descreveremos a seguir, pois, a partir das observações e análises preliminares do corpus de estudo, verificamos a necessidade de sua expansão e buscamos uma metodologia que nos pareceu mais adequada para atingir nossos objetivos de descrição de letras de música, preservando a natureza teórica da pesquisa (Linguística de Corpus): a análise multidimensional.

Neste capítulo, descrevemos o estudo piloto que forneceu subsídios para a nossa investigação final. O capítulo seguinte descreve os passos metodológicos de nosso estudo principal, seguido por nossa análise, discussão e considerações finais.

#### 4. Metodologia de Pesquisa do Estudo Principal

No capítulo anterior, descrevemos nosso estudo piloto através do qual investigamos convergências lexicais entre letras de música e a língua inglesa em geral. Verificamos que as letras de música têm muita semelhança com o inglês usado em outras manifestações da língua, sejam escritas ou faladas. Tais constatações despertaram a necessidade de descrever nosso corpus de letras de música com mais detalhe e, conseqüentemente, buscar um novo procedimento de análise.

Partimos da ideia de que a variação de estilos musicais pode ser explicada, em parte, pela variação linguística das letras não por meio de uma oposição binária, ou seja, ou é *country* ou é *pop*; ou é *pop* ou é *rock*, etc. Acreditamos que as letras de música devem apresentar semelhanças e diferenças, quando analisadas por suas características linguísticas, compondo diferentes dimensões de variação. Ou seja, segundo critérios linguísticos, muitas músicas *country* podem ser mais ou menos próximas das de *pop*, enquanto muitas de *country* podem ser bastante diferentes das de *pop* e assim por diante.

O objetivo principal desta pesquisa é, portanto, desvendar as dimensões de variação linguística que permitem uma caracterização da música popular anglo-americana a partir da observação de suas letras. Para a sua realização apoiamos-nos teoricamente na Linguística de corpus, cujos princípios de análise, como frequência, listas de palavras, pacotes lexicais e, sobretudo a metodologia de análise multidimensional (AMD), conforme proposta por Biber (1988), nortearam nossos procedimentos e decisões metodológicas. Face a esse objetivo, orientamo-nos pelas seguintes questões já mencionadas:

1. Quais os principais fatores<sup>129</sup> responsáveis pela variação linguística presente nas letras de música popular anglo-americana?
2. Quais dimensões (funções comunicativas) expressam os fatores obtidos?
3. Como essas dimensões são representadas em relação aos diferentes gêneros musicais presentes em nosso corpus de estudo?

---

<sup>129</sup> Agrupamentos de variáveis (características linguísticas) coocorrentes, como dissemos anteriormente.

4. Como essas dimensões são representadas em relação aos diferentes artistas selecionados para o estudo?
5. Como essas dimensões são representadas ao longo do tempo por nós estudado (1940-2009)?

Essas questões estão amparadas especialmente pela visão da língua como sistema probabilístico (BERBER SARDINHA, 2004a), o qual assume que embora muitas escolhas e combinações lexicais sejam possíveis, elas não ocorrem da mesma forma nem com a mesma frequência (FIRTH, 1957; HALLIDAY; HASAN, 1989; HALLIDAY; 1991; SINCLAIR; 1991; HALLIDAY; WEBSTER; 2002). Essa noção de que a variação de escolhas linguísticas não ocorre aleatoriamente, mas segue certa padronização, subjaz à ideia de que a observação lexical empírica vai permitir uma descrição mais completa e mais precisa da língua, pois evidenciará aquilo que é mais típico em um determinado contexto.

A fim de respondermos a essas questões, desenvolveremos uma análise seguindo o modelo proposto por Douglas Biber (BIBER, 1988 et seq.) para uma Análise Multidimensional, o qual também assume que as características linguísticas podem ser descritas a partir da observação de suas associações e frequência nos textos, de acordo com o contexto, por meio de procedimentos estatísticos, cujos resultados são interpretados qualitativamente. As etapas para a AMD propostas por Biber (1988) e resumidas por Berber Sardinha (2000b; 2004, p. 305-306), conforme apresentado no capítulo teórico, são as seguintes:

- Levantamento das características linguísticas relevantes por meio de ampla consulta à literatura disponível.
- Coleta ou adoção de um corpus de dados linguísticos representativo e compatível com as metas da análise.
- Transformação das características linguísticas em variáveis quantificáveis.
- Codificação dos dados baseada nas variáveis selecionadas, usando ferramentas computacionais para análise automática, semiautomática (interativa) ou manual.
- Conferência manual da codificação feita por computador para checar sua exatidão.
- Padronização das frequências (em geral por 1.000 palavras), para permitir a comparação entre variedades (textos, registros ou corpora) de extensões

diferentes.

- Computação de frequências médias de cada variável.
- Análise Fatorial inicial, a fim de obter os pesos (*loadings*) de cada variável em cada variedade.
- Determinação do número de fatores por meio da aplicação de técnicas como observação dos valores *eigen* (*eigenvalues*) em um gráfico *scree* (*scree plot*).
- Análise Fatorial posterior, com a rotação dos fatores.
- Cálculo de escores de cada texto por fator pela padronização dos escores com base na média e no desvio padrão.
- Cálculo de escores médios de cada variedade por fator.
- Interpretação de cada fator e rotulação das dimensões.

Neste capítulo apresentaremos a descrição de cada uma dessas etapas metodológicas, da maneira que foram desenvolvidas para nossa pesquisa. Primeiramente, faremos a descrição dos corpora utilizados no estudo: O corpus Google N-Grams e nosso corpus de estudo, de letras de música. Gostaríamos de ressaltar, contudo, que, embora num determinado momento o corpus tenha sido estabelecido como final – como foi utilizado no momento da análise –, nosso corpus de estudo está sempre em crescimento, pois consideramos que a língua da música, assim como a língua em geral, está em constante crescimento e movimento e portanto, não pode ser considerado como finita.

Em seguida, descreveremos as características linguísticas (partes do discurso, agrupamentos semânticos e pacotes lexicais) selecionadas para nossa análise e sua transformação em variáveis quantificáveis. Depois disso, descreveremos os procedimentos estatísticos que nos levaram à extração de fatores (análise fatorial) de frequência de características linguísticas coocorrentes, cuja interpretação gerou as dimensões que procurávamos.

Faz-se necessário ainda um esclarecimento adicional em relação ao conteúdo deste capítulo no que diz respeito aos procedimentos metodológicos. Muitas de nossas decisões metodológicas foram tomadas a partir dos resultados obtidos em cada uma das etapas da análise estatística. Dessa forma, exploraremos todas essas etapas metodológicas neste capítulo, relegando ao próximo a interpretação dos fatores.

## 4.1. Os Corpora utilizados

### 4.1.1. O corpus de estudo

Nesta seção descrevemos nosso corpus de estudo, que chamaremos de CM (corpus de música), e nossos procedimentos de coleta.

O CM é uma complementação e adequação do corpus que coletamos e foi usado em nosso estudo piloto, descrito no capítulo anterior. Em outras palavras, o corpus é formado por uma compilação de mais de 6.000 letras de música compostas e gravadas por artistas falantes nativos de língua inglesa, mais especificamente, americanos, britânicos e canadenses, conforme explicitaremos a seguir.

Numericamente, o CM no momento de nossa análise apresentava a seguinte descrição:

TOTAL	letras de música	6.290
	Ocorrências	1.151.280
	Formas	23.355

quadro 12: descrição do corpus principal CM (corpus de música)

Como se pode observar, nosso corpus é composto por 6.290 letras de música (ou canções, como também nos referimos a elas) que juntas somam 1.151.280 palavras corridas (ocorrências ou *tokens*), que são resultados do uso de 23.355 palavras diferentes (formas ou *types*).

Um corpus com mais de um milhão de palavras é considerado um corpus médio-grande (BERBER SARDINHA, 2004, p. 26). É importante ressaltar a importância do tamanho do corpus, ou seja, sua representatividade. O tamanho de um corpus deve levar em conta o que se pretende estudar e a especialização do corpus. Para um corpus especializado, ou seja, que foi coletado especialmente para atender aos critérios de pesquisa, como é o nosso, o autor sugere um tamanho mínimo de 91.161 palavras (BERBER SARDINHA, 2004, p. 98)<sup>130</sup>. Dessa forma, em relação ao número de

<sup>130</sup> Para a obtenção do valor mínimo de palavras o autor aplicou uma metodologia de estimação de amostras linguísticas significativas, sugerida por Biber (1990, 1993), cf, BERBER SARDINHA, 2004, p. 98)



palavras, constatamos a adequação de nosso corpus de estudo.

Em relação à origem dos dados, nosso corpus é caracterizado por aquilo que chamamos de “conjunto da obra” de determinado artista. Ou seja, selecionamos artistas que tiveram músicas nas paradas de sucesso (cf. STARR; WATERMAN, 2007 e informações dos *Charts*, obtidas através dos sites <http://www.billboard.com> e <http://www.musicimprint.com/Chart.aspx?id=C000002>), de maneira (praticamente) aleatória<sup>131</sup>.

No que diz respeito ao período de tempo, iniciamos nossa coleta levando em consideração os anos entre 1950 e 2000, por compreender um período de cinquenta anos. A seleção por artista, implicou que considerássemos também composições/gravações realizadas antes de 1950, as quais agrupamos para a década de 1940 a 1949, e gravações/composições realizadas após 2000. Neste último caso, levando em consideração que vislumbramos a utilização dos dados para futura prática pedagógica e a indicação (advinda de nossa experiência profissional) de que aprendizes preferem ouvir “a música do momento”, selecionamos também bandas pertencentes exclusivamente à década que compreende os anos entre 2000 e 2009.

Feitas essas considerações, apresentamos a composição do corpus de estudo no quadro a seguir, no qual também aparece o número de palavras presentes nas letras de músicas por artista:

ARTISTA	LETRAS DE MÚSICA	TOKENS
1. Aerosmith	152	34.385
2. Beach Boys	276	58.910
3. Beatles	281	46.911
4. Bon Jovi	127	33.205
5. Creed	60	10.614
6. Def Leppard	142	31.939
7. Elton John	325	53.469
8. Elvis Presley	305	47.359

<sup>131</sup> Embora tenhamos considerado artistas que estiveram presentes nas paradas várias vezes em anos diferentes e artistas que tiveram suas músicas entre as 100 músicas mais famosas de todos os tempos (*top songs of all time*, cf: <http://www.musicimprint.com/Chart.aspx?id=C000002>), como Beatles; Rolling Stones; Elvis Presley; Led Zeppelin; Queen; Nirvana; Johnny Cash; U2; Madonna; Michael Jackson; Ray Charles; Frank Sinatra e os Beach Boys.

9. Foo Fighters	109	17.879
10. Frank Sinatra	1.218	135.747
11. Greenday	159	29.110
12. Iron Maiden	215	45.163
13. Johnny Cash	439	96.508
14. The Killers	51	12.200
15. Led Zeppelin	79	15.711
16. Lenny Kravitz	125	25.001
17. Madonna	167	45.973
18. My Chemical Romance	61	15.237
19. Metallica	119	26.221
20. Michael Jackson	99	28.008
21. Nickelback	79	20.152
22. Nirvana	98	14.832
23. Paramore	50	11.148
24. Pearl Jam	85	13.819
25. Pink Floyd	163	27.307
26. Queen	172	41.525
27. Queensryche	112	21.202
28. Ray Charles	206	36.524
29. Red Hot Chilly Peppers	169	38.909
30. Rolling Stones	430	79.985
31. Simple Plan	49	12.636
32. U2	168	23.681
TOTAL	6.290	1.151.280

quadro 13: descrição detalhada do corpus de estudo (CM)

Foram coletadas, portanto, 6.290 músicas, compostas e/ou gravadas por 32 artistas diferentes, desde 1940 até 2009. Em relação ao corpus de estudo utilizado na pesquisa piloto, podem ser observadas algumas alterações. Isso ocorre porque alguns artistas tiveram músicas incluídas (por não terem sido coletadas anteriormente, e.g. Aerosmith) e outros tiveram músicas retiradas, a fim de se evitar duplicidade no corpus total (por aparecerem gravadas também por outros artistas, e.g. Frank Sinatra).

Além disso, considerando que, para o estudo principal, estávamos focalizando nossa análise em diferentes estilos musicais ao longo do tempo, excluímos o subcorpus das músicas mais tocadas na MTV em agosto de 2008, que havíamos chamado de TOP 100, com 121 músicas e o subcorpus das músicas compostas para o público adolescentes, Teen, com 202 músicas. Essa decisão foi tomada levando em conta a

caracterização geral do corpus, pois todos os outros dados coletados foram organizados por banda e por década. Como esses dois subcorpora não se enquadravam nessas características (as letras de música haviam sido compostas por variados artistas), eles foram descartados para esta análise. Por outro lado, incluímos o conjunto da obra de outros artistas, a saber: Beach Boys; Elvis Presley; Johnny Cash e Michael Jackson, que forneceram maior variedade temporal e estilística.

Os dados foram organizados da seguinte maneira: As letras de músicas foram agrupadas primeiramente por década: de 1940 a 1949; de 1950 a 1959; de 1960 a 1969; de 1970 a 1979, e assim por diante, até de 2000 a 2009. Cada década recebeu uma subdivisão por artista e cada artista teve suas músicas arquivadas pelo ano de gravação. Melhor visualizado nos quadros que seguem:

CM						
1940-1949	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2009
Frank Sinatra	Elvis Presley Frank Sinatra Johnny Cash Ray Charles	Beach Boys Beatles Elton John Elvis Presley Frank Sinatra Johnny Cash Led Zeppelin Pink Floyd Ray Charles Rolling Stones	Aerosmith Beach Boys Beatles Elton John Elvis Presley Frank Sinatra Johnny Cash Led Zeppelin Michael Jackson Pink Floyd Queen Ray Charles Rolling Stones	Aerosmith Beach Boys Bon Jovi Def Leppard Elton John Frank Sinatra Iron Maiden Johnny Cash Led Zeppelin Lenny Kravitz Madonna Metallica Michael Jackson Nirvana Pink Floyd Queen Queensryche Ray Charles Red Hot Chilly peppers Rolling Stones U2	Aerosmith Beach Boys Bon Jovi Creed Def Leppard Elton John Foo Fighters Frank Sinatra Green Day Iron Maiden Johnny Cash Led Zeppelin Lenny Kravitz Madonna Metallica Michael Jackson Nirvana Pearl Jam Pink Floyd Queen Queensryche Ray Charles Red Hot Chilly peppers Rolling Stones U2	Aerosmith Bon Jovi Creed Def Leppard Elton John Foo Fighters Green Day Iron Maiden Killers Lenny Kravitz Madonna Metallica Michael Jackson My Chemical Romance Nickelback Nirvana Paramore Pearl Jam Pink Floyd Queen Queensryche Ray Charles Red Hot Chilly peppers Rolling Stones Simple Plan U2

quadro 14: Corpus de Música (CM) subdividido por década

O quadro anterior mostra nosso corpus de estudo com sua subdivisão por década, por artista. Cada artista ainda teve suas músicas arquivadas por ano de gravação dentro de cada década. A título de ilustração, apresentamos o ano de 1966 para o cantor Elvis Presley na figura disposta a seguir:

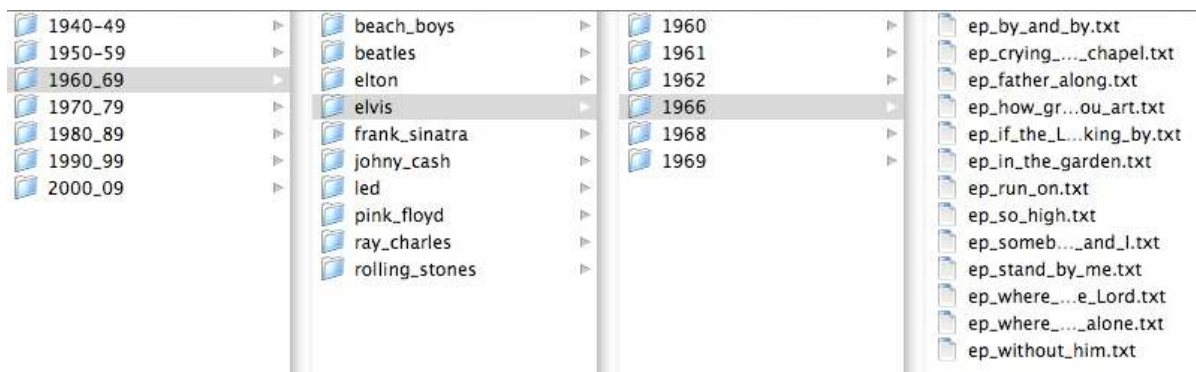


figura 1: CM, na década de 1960, com as músicas gravadas por Elvis Presley em 1966

Na primeira coluna estão as décadas que fazem parte de nosso estudo (1940-49; 1950-59; 1960-69; 1970-79; 1980-89; 1990-2000 e 2000-09). Na segunda coluna aparecem os artistas que tiveram gravações entre os anos de 1960 e 1969, selecionado na primeira. O cantor Elvis Presley está selecionado na segunda coluna. Para esse período, o cantor teve músicas gravadas nos anos de 1960, 1961, 1962, 1966, 1968 e 1969. Os anos ausentes representam anos em que o artista não lançou nenhum disco. O ano de 1966, selecionado na terceira coluna, apresenta, na quarta coluna, a lista das 13 músicas gravadas por aquele artista nesse ano.

Esse procedimento foi adotado para todas as décadas e todos os artistas presentes em nosso corpus de estudo. Dada a grande quantidade de dados, 6.290 músicas, consideramos extremamente extenso apresentar quadros para cada artista, para cada período de tempo, assim, apresentamos apenas uma amostra, na figura 1.

Outro aspecto interessante é que enquanto alguns artistas gravaram e lançaram discos dentro de maiores intervalos de tempo, alguns, como Frank Sinatra, por exemplo, chegaram a lançar três discos no mesmo ano. Esse comportamento de gravação nos levou a optar pela não subdivisão por discos, mas sim por anos e, especialmente, por décadas.

Quando pensamos em nossos dados em termos de quantidade de arquivos (mais de 6.000 letras de música), nosso corpus parece imenso, especialmente se comparado ao de

Biber (1988), por exemplo, que para sua análise multidimensional textual da língua inglesa usou 481 textos, ou o de Kauffmann (2005), que para sua AMD de gêneros jornalísticos usou 1.431. Todavia, é importante observar que os textos coletados por esses autores contêm cerca de 200 a 500 palavras cada um, o que levou Kauffman a um corpus com aproximadamente 500 mil palavras e Biber a um corpus com pouco mais de 1 milhão de palavras. As letras de música por sua vez correspondem a pequenos textos, muitos deles com menos de 100 palavras. Como objetivamos atingir um grande número de palavras, conseqüentemente, temos um número muito maior de textos/arquivos a serem analisados.

#### **4.1.2. Corpus Google N-Gram**

Outro corpus utilizado em nossa pesquisa é o corpus do Google N-Gram. Esse corpus contém palavras em inglês agrupadas em pacotes lexicais, acompanhadas por sua contagem de frequência. Os dados foram gerados a partir de aproximadamente 1 trilhão de palavras (*tokens*), advindas de páginas de acesso público na *web*, através do buscador Google e disponibilizados para estudos e pesquisa por meio de um consórcio aberto entre universidades, companhias e laboratórios de pesquisa governamentais, o LDC (*Linguistic Data Consortium*), que está hospedado na Universidade da Pensilvânia. (cf. BRANTS; FRANZ, 2006)

Um n-grama é uma sequência de um número definido de palavras (ou letras), se a sequência for de duas palavras temos bigramas, se for de três palavras temos trigramas. Como o próprio nome diz, os dados do corpus Google N-Grams são apresentados de acordo com a extensão dos n-gramas, que vai de unigramas (palavras isoladas) a cinco palavras juntas. O objetivo da confecção e disponibilização do corpus é que esses dados sejam usados para modelagem estatística da língua, como, por exemplo, para a tradução automática ou reconhecimento de discurso, assim como para outros usos.

Consideramos os trigramas, que somam 977.069.902, como os dispostos a seguir:

ceramics collectables collectibles	55
ceramics collectables fine	130
ceramics collected by	52
ceramics collectible pottery	50
ceramics collectibles cooking	45

Para esta pesquisa o corpus Google N-Grams foi utilizado da seguinte maneira: Primeiramente, foi feita a extração de todos os trigramas presentes no corpus de estudo, dos quais selecionamos os 200 mais frequentes, tais como “*I love you*”, que ocorreu 630 vezes; “*I want you*”, 550 vezes; “*oh oh oh*”, 499 vezes; “*I don’t know*”, 382 vezes; e “*I don’t want*”, 259 vezes. Em seguida, a frequência desses trigramas foi verificada no corpus do Google N-Grams, para cada letra de música, por meio de uma rotina desenvolvida pelo professor orientador, em linguagem SQL e Shell. Dessa forma, conseguimos detectar em quais canções aparecem os trigramas mais frequentes no corpus de estudo e sua frequência nas letras de música e no inglês geral de internete.

Descritos o corpus de música e suas subdivisões e o corpus do google e sua utilização na pesquisa, passaremos a apresentar seleção de características linguísticas consideradas para o nosso estudo e, conseqüentemente, as variáveis utilizadas na análise fatorial.

#### **4.2. Características Linguísticas e as variáveis da pesquisa**

No início deste capítulo e anteriormente no capítulo teórico apresentamos os procedimentos metodológicos para a AMD, conforme propostos por Biber (1988). Segundo o autor, a primeira etapa é o levantamento das características linguísticas, seguida pela transformação das características selecionadas em variáveis quantificáveis. Para que isso seja possível, considerando a dimensão dos dados, é feita uma codificação dos dados usando ferramentas de análise automática, ou etiquetagem, que deve ser conferida manualmente.

Características linguísticas, *traços linguísticos*, representam os elementos lexicais e gramaticais típicos de um texto, sendo representadas pelas classes gramaticais (substantivos, adjetivos, conjunções, etc.). Para que os dados linguísticos possam ser contabilizados, o primeiro passo, após a coleta do corpus, é a sua etiquetagem morfossintática (ou *POS tagging*<sup>132</sup>).

A etiquetagem do nosso corpus de estudo foi feita automaticamente em termos morfossintáticos e semânticos. “O etiquetador serve para inserir automaticamente, no corpus, códigos que indicam a classe gramatical de cada palavra” (BERBER SARDINHA, 2004a, p.113) e permite que grandes quantidades de dados sejam analisadas.

Primeiramente, nosso corpus foi submetido ao etiquetador *Tree-Tagger*, que é um etiquetador morfossintático automático “preparado para etiquetar em inglês, usando o esquema da Universidade da Pensilvânia”<sup>133</sup> (BERBER SARDINHA, 2004a, p.114), disponível no sítio do CEPRIL ([www2.lael.pucsp.br/corpora](http://www2.lael.pucsp.br/corpora)). A versão usada para esta pesquisa, devido à quantidade grande de arquivos a ser etiquetado um a um, não foi *online*, mas uma *offline*, acionada por um *script* especialmente desenvolvido pelo professor orientador. Esse etiquetador atribui uma etiqueta morfossintática e um lema<sup>134</sup> a cada palavra de cada música, automaticamente. Os lemas não foram usados nesta pesquisa. As 48 etiquetas consideradas para o nosso estudo, acompanhadas de suas representações/explicações, podem ser vistas a seguir:

CC = conjunção coordenativa	POS = terminação possessiva ('s)	VH = verbo to have
CD = números cardinais	PP_ = pronome possessivo	VHD = verbo to have no passado
DT = determinantes	PP = pronome pessoal (mine, yours...)	VHG = verbo to have sem flexão
EX = existencial (there)	RB = advérbio	VHN = verbo to have no particípio passado
FW = palavra estrangeira	RBR = advérbio comparativo	VHP = verbo to have no presente plural
IN = preposição ou conjunção subordinativa	RBS = advérbio superlativo	VHZ = verbo to have no presente singular
JJ = adjetivos	RP = partícula	VV = verbo
JJR = comparativos	SYM = símbolo	VVD = verbo no passado
JJS = superlativos	TO = preposição TO	
LS = lista (marcadores como	UH = interjeição	

<sup>132</sup> Do inglês Parts-of-Speech tagging

<sup>133</sup> Ver <http://www.cis.upenn.edu/~treebank/>

<sup>134</sup> Um lema é a representação de uma palavra em sua forma básica, correspondendo a todas as suas formas flexionadas. Por exemplo, o lema do verbo “ser” corresponde a “é, são, era, foi, foram, estávamos” etc.

números e letras)	VB = verbo to be no infinitivo	VVG = verbo no gerúndio
MD = modalizadores (verbos como should, may, might...)	VBG = verbo to be no gerúndio	VVN = verbo no particípio
NN = substantivo singular	VBD = verbo to be no passado	VVP = verbo no presente plural
NNS = substantivo plural	VBN = verbo to be no particípio passado	VVZ = verbo no presente 3 <sup>a</sup> . Pessoa
NP = substantivo próprio	VBP = verbo to be no presente	WDT = determinante (WH); pronomes relativos
NPS = substantivo próprio plural	VBZ = verbo to be 3 <sup>a</sup> . Pessoa singular	WP_ = (WH) pronome possessivo whose
PDT = pré-determinante		WP = pronome WH
		WRB = advérbio WH

quadro 15: etiquetas morfossintáticas do *Penn-tree tagger*

A título de ilustração, reproduzimos os três primeiros versos da canção “*Empty garden*”, gravada em 1982 pelo artista britânico Elton John, antes do processo de etiquetagem:

“What happened here  
As the New York sunset disappeared  
I found an empty garden among the flagstones there”

Após o processo de etiquetagem, cada palavra fica posicionada em uma única linha e aparece acompanhada de sua etiqueta (no centro da listagem) e de seu lema (palavra de base), na coluna à direita, conforme pode ser visto a seguir:

What	WP	What
happened	VVD	happen
here	RB	here
As	IN	as
the	DT	the
New	NP	New
York	NP	York
sunset	NN	sunset
disappeared	VVD	disappear
I	PP	I
found	VVD	find
an	DT	an
empty	JJ	empty
garden	NN	garden
among	IN	among
the	DT	the
flagstones	NNS	flagstone
there	RB	there

Como vemos, cada palavra está categorizada por sua classe gramatical de acordo



com as etiquetas listadas no quadro anterior: ‘*what*’, pronome *wh*; ‘*happened*’, verbo no passado; ‘*here*’, advérbio; etc.

Levando em conta a extensão de nossos dados, mais de 1 milhão de palavras, e a margem de erro mínima do etiquetador, da ordem de cinco por cento (RAMILO; FREITAS, 2002), a etiquetagem morfossintática não foi conferida manualmente em sua totalidade.

Em segundo lugar, nosso corpus de estudo foi também etiquetado por suas características semânticas, através do programa *semantic tagger*, criado pelo professor orientador. Esse programa foi baseado em outros dois já existentes: *wsd.pl* e o SUMO.

O *wsd.p* (PEDERSEN; MICHELIZZI, 2004-2008)<sup>135</sup> faz dezambiguação de sentido de palavras, por exemplo: nas expressões “*the river bank*” e “*the bank of Boston*” a palavra ‘*bank*’ é ambígua para o computador (não para o ser humano), porque pode significar tanto ‘margem do rio’ quanto ‘instituição financeira’, o programa consegue diferenciar o sentido das duas.

O outro programa é o SUMO<sup>136</sup>, que é uma ontologia qualitativa modular do conhecimento, isto é, contém categorias gerais do conhecimento humano, divididas por área. Em outras palavras, substantivos, verbos, adjetivos e advérbios foram agrupados em conjuntos de palavras que se relacionam semanticamente<sup>137</sup>. Esses conjuntos estão ligados por suas relações semânticas e lexicais. A palavra “*table*”, por exemplo, segundo o programa<sup>138</sup>, está inclusa em cinco diferentes categorias: “*Table*”, quando apresenta o sentido de peça de mobília com tampo plano e quatro pés, utilizada para comer; “*Food*”, quando se refere à comida (e.g. “*she sets a fine table*”, ou em português, “ela prepara uma boa mesa”); “*ContentBearingObject*”, ou objeto de conteúdo, quando se refere a um conjunto de dados organizados em colunas e linhas (tabela de dados); “*Meeting*”, quando se refere às pessoas sentadas à mesa para reunião; e “*Mesa*”, quando se refere a acidente geográfico, terra alta.

O *semantic tagger* desambigua cada palavra das músicas com o *wsd.pl*; em

<sup>135</sup> Ver: <http://search.cpan.org/perldoc?wsd.pl>

<sup>136</sup> SUMO equivale a Suggested Upper Merged Ontology. Ver: <http://www.ontologyportal.org/> - ontologia qualitativa

<sup>137</sup> conforme <http://wordnet.princeton.edu/>

<sup>138</sup> ver: <http://sigma.ontologyportal.org:4010/sigma/WordNet.jsp?word=table&POS=1>

seguida, a partir do arquivo, o programa busca as palavras no SUMO para encontrar as categorias referentes a ela. Vejamos um trecho de um arquivo processado pelo *semantic tagger* contendo os mesmos três versos da mesma canção apresentada anteriormente, “*Empty garden*”, após a etiquetagem semântica:

```

1,What#NT,No Tag,\N
2,happen,VERB,Process
3,here,ADV,TemporalRelation
4,as#NT,No Tag,\N
5,the#NT,No Tag,\N
6,New,NOUN,\N
7,York,NOUN,FamilyGroup
8,sunset,NOUN,Sunset
9,disappear,VERB,located
10,I#NT,No Tag,\N
11,find,VERB, Process
12,an#NT,No Tag,\N
13,empty,ADJ,SubjectiveAssessmentAttribute
14,garden,NOUN,CultivatedLandArea
15,among#NT,No Tag,\N
16,the#NT,No Tag,\N
17,flagstone,NOUN,CorpuscularObject

```

Palavras como “*what*”, “*as*”, “*the*”, “*I*”, “*an*” e “*among*” não recebem etiqueta (NT = no tag), porque não pertencem às categorias semânticas contidas no programa. Palavras como “*happen*” e “*find*” são verbos e foram categorizadas como processo; “*here*” pertence à família de relações temporais e assim por diante. O objetivo da etiquetagem semântica é obter um reflexo dos assuntos tratados nas letras das músicas.

As associações semânticas verificadas nas músicas foram agrupadas pelo programa *semantic tagger* sob 100 diferentes títulos: agente; aparelho (*device*); aprender; área de terra (*land area*); artefato; atributo subjetivo; aumento; caminhar; caminho; cantar; capacidade (*capability*); casa; causas; cérebro (*brain*); cidades; classificação; cobertura (*covering*); colocar; comunicação; crenças (*believes*); cultivo de plantas; dança; declarar; desejos; dia; diminuição; direito (*confers right*); e (*and*); entre (*between*); estado emocional; estado ou província; exemplo; expectativa; falar; feminino; fonte; grupo de pessoas; habitar; homem; humano; *I* (eu); igualdade; impacto; infelicidade; ingestão; interação social; interpretar; intervalo de tempo; jogos; lembrar; localizado; luz radiante;

maior (*greater than*); mais cedo (*earlier*); manter; mão (*hand*); masculino; medida de tempo; modo; morte; movimento; movimento físico (*body motion*); música; não (*not*); necessidade; objeto; objeto corpuscular (*corpuscular object*); objeto de conteúdo (*content bearing object*); obrigação; obter (*getting*); oferecer (*giving*); olhar (*looking*); órgão (*organ*); papel social; para todos (*for all*); parte; partir (*leaving*); passado finito; perseguir; posição; positividade; processo; processo educacional; processo intencional; processo psicológico; proposição; que envolve (*entails*); querer.; raciocinar; região; relação temporal; remover; saber; selecionar; solicitar; som radiante; texto; transferência; transporte; verdade.

A categorização semântica foi realizada automaticamente e viabilizada por meio de um formulário gerado por um script desenvolvido especialmente para esta pesquisa. As palavras agrupadas em cada categoria foram conferidas manualmente. Assim, pudemos selecionar as categorias que julgamos apresentar maior consistência. A seguir, apresentamos um exemplo de uma categoria que incluímos e de uma categoria que descartamos.

<b>categoria descartada: Agent (agente)</b>	<b>categoria mantida: Day (dia)</b>
===== Agent ===== >>> Keep this category? yes no 716 God 303 Cause 128 god 77 cause 27 grace 7 immortal 5 trinity 5 steamroller 5 Grace 4 cart 3 Killer 3 jockey 2 spectre 2 divinity 2 deity 1 spirit	===== Day ===== >>> Keep this category? yes no 5072 now 3231 day 255 Christmas 246 tomorrow 188 today 187 yesterday 53 Day 42 daylight 33 Days 31 holiday 28 birthday 28 Holiday 7 Thanksgiving 5 payday 4 picnic 4 jubilee

1 specter	4 nowadays
1 phantom	4 Birthday
1 nature	2 Tomorrow
1 Jehovah	2 morrow
1 incarnate	2 midsummer
1 Cart	1 solstice
	1 Pentecost
	1 Yesterday
	1 Halloween
	1 weekday

quadro 16: amostra de categorias semânticas (agente e dia) extraídas pelo programa *semantic tagger*

No quadro 16 observamos as categorias “agente” e “dia”, extraídas pelo programa *semantic tagger*. A primeira linha apresenta o título da categoria, a segunda apresenta um questionamento sobre incluir ou não a categoria na análise. Em seguida, estão as palavras antecidas por sua frequência, conforme foram consideradas para a categoria em questão. As palavras agrupadas sob o título agente (*agent*) apresentaram certa inconsistência, consideramos que as palavras “*Cause*” e “*cause*”, cujas ocorrências são altas, 303 e 77 vezes; assim como “*grace*”; “*immortal*”; “*spectre*”; “*nature*” e “*incarnate*” não podem ser consideradas como agente, especialmente quando contrastadas com os outros itens categorizados da mesma forma, como “*God*”; “*Trinity*” ou “*killier*”. Portanto, optamos por não considerar essa categoria. Já a categoria “dia”, apresentou uma seleção mais pertinente de palavras que se relacionam semanticamente com o título, tais como: “*now*”; “*day*”; “*Christmas*”; “*tomorrow*”; “*yesterday*”; “*birthday*”, “*solstice*”, etc.

Seguindo esse padrão de conferência, foram incluídas em nossa análise 39 categorias semânticas, sob os seguintes títulos (traduzidos para o português):

artefato	processo intencional	solicitar
movimento físico (body motion)	processo psicológico intencional	cantar
capacidade (capability)	masculino	papel social
dança	movimento	falar
dia	música	declarar
morte	necessidade	atributo subjetivo
estado emocional	não (not)	intervalo de tempo
igualdade	passado finito	medida de tempo
feminino	processo	transporte
grupo de pessoas	proposição	verdade
casa	perseguir	infelicidade
humano	luz radiante	caminhar
impacto	som radiante	querer

quadro 17: 39 categorias semânticas incluídas em nossa análise

Além dessas categorias semânticas que selecionamos a partir daquelas geradas automaticamente pelo *semantic tagger*, criamos manualmente mais seis categorias que nos chamaram a atenção pela persistência de palavras que se referiam a esses temas no corpus: religião; sexo; amor; violência; drogas e fantástico (referências a seres e objetos sobrenaturais). Sob o título religião, por exemplo, incluímos palavras como *Abraham, Adam, Allelujah, Angel, Armageddon, Augustin, baptized, blessing, Buddah, cahedral, celestial, chruch, damnation, demon, devil, eden, devine, faithfull* e assim por diante. Para sexo foram incluídas palavras como *attraction, bimbo, chamber, lust, excited, erotic, dates, lover, masturbation, porn, etc.* Na categoria amor, incluímos *love, dream, affection, affair, bridal, cheat, darling, hate, husband, kiss*, entre outras. Sob o título violência categorizamos palavras como *accident, annihilation, assassination, battle, bomb, cannon, fight, gun, gunshot, Hitler, kill, outlaw, etc.* Como drogas categorizamos palavras como *marijuana, alcohol, cocaine, novocaine, prozac, drugs, inject, ellation, halluciantion, junk, LSD*, entre outras. Finalmente, sob o título fantástico encontram-se palavras como *Alladin, alien, Apollo, dragon, Camelot, Dracula, Elf, Gnome, Hercules, Icarus, Mermaid, mojo, mystic, superman, sandman, tarot, vampire, zombie, zodyac*, etc. Essas categorias semânticas ‘a posteriori’ foram criadas pela observação de liestas de frequência criadas pelo computador, como explicamos no capítulo 3.

Além das categorias que acabamos de descrever, 48 categorias morfossintáticas e 45 categorias semânticas (39 extraídas automaticamente, somadas às 6 categorias criadas manualmente), julgamos pertinente verificar o comportamento de pacotes lexicais, também chamados de *lexical bundles*. Pesquisas como a de Crossley e Lowerse (2007) demonstraram a importância de se localizar colocações (palavras que aparecem próximas umas das outras com frequência), ao aplicarem a análise fatorial (AMD) com pacotes lexicais, contrastando textos falados e escritos. Segundo os autores “sequências de múltiplas palavras são características linguísticas importantes na construção de funções discursivas e comunicativas”<sup>139</sup> (CROSSLEY; LOWERSE, 2007, p. 275).

Sendo assim, atribuímos mais duas categorias considerando pacotes lexicais de três palavras, ou trigramas, como explicamos anteriormente, que chamamos de *song bundles* e *google counts*. A primeira representa a contagem de quantas vezes aparece em cada música qualquer um dos 200 trigramas mais frequentes do nosso corpus de estudo. O objetivo dessa contagem é medir a convencionalidade das músicas, pois quanto maior quantidade desses *bundles* aparecerem numa mesma letra de música mais convencional ela será, já que repete partes encontradas comumente em outras letras. Isso é o mesmo que dizer que uma música com muitos *bundles* está usando partes de outras músicas. A segunda, por sua vez, calcula a frequência de cada um desses trigramas na internete (corpus Google N-Gram). Quando verificamos que os trigramas que ocorrem nas letras de música também ocorrem no inglês geral, temos indicação da proximidade (ou distância) das letras de música com a língua geral (no canal online, neste caso).

Em suma, nossa análise contou com a seleção de 93 características linguísticas, ou seja, as 48 categorias morfossintáticas e as 45 categorias semânticas. A essas características foram acrescentadas a contagem de trigramas no corpus de letras de música e no corpus do google N-Grams – que chamamos respectivamente de *song bundles* e *google counts* – e a contagem de palavras individuais (*types*) e ocorrências (*tokens*) do corpus de estudo. Essas características, quando consideradas na análise fatorial, representam as **97 variáveis** iniciais de nossa pesquisa.

Os dados foram arquivados no formato de planilha do Excel, vez que o programa

---

<sup>139</sup> Do original: “multi-word sequences are important linguistic features in the building of discourse and communicative functions”

estatístico de análise fatorial os computa a partir dessa formatação. Nossa planilha, portanto, ficou com 98 colunas (as 97 variáveis, mais uma coluna com os nomes dos arquivos) e 6.291 linhas (6.290 músicas/arquivos mais a primeira linha com os títulos).

Devido à extensão da planilha, optamos por não reproduzi-la em sua totalidade. Uma amostra da planilha pode ser visualizada no ANEXO 2. Porém, a título de ilustração, reproduzimos parte dela a seguir:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	file	tokens	types	CC	CD	DT	EX	FW	IN	JJ	JJR	JJS	LS	MD	NN	NNS
2	1940_fs_April_played_a_fiddle.txt.token.tag	85	53	9	0	10	0	0	4	2	0	0	0	1	16	2
3	1940_fs_Empty_saddles.txt.token.tag	107	56	1	0	10	0	0	13	9	0	0	0	1	15	7
4	1940_fs_Every_night_about_this_time.txt.token.tag	124	50	3	0	21	0	0	16	3	0	0	0	0	24	4
5	1940_fs_Hands_across_the_table.txt.token.tag	43	35	1	0	3	0	0	6	2	0	0	0	0	4	4
6	1940_fs_Head_on_my_pillow.txt.token.tag	78	50	2	0	2	0	0	12	3	0	0	0	0	13	4
7	1940_fs_ILL_Take_talulah.txt.token.tag	124	73	1	1	13	0	0	4	1	0	0	0	5	16	2
8	1940_fs_ILL_never_let_a_day_pass_by.txt.token.tag	90	61	5	1	10	0	0	12	4	0	0	0	2	16	0
9	1940_fs_ILL_never_smile_again.txt.token.tag	89	42	0	0	1	0	0	14	4	0	0	0	5	8	2
10	1940_fs_ILL_string_along_with_you.txt.token.tag	98	56	4	2	8	0	0	19	7	0	0	0	2	13	3
11	1940_fs_I_could_make_you_care.txt.token.tag	78	46	0	0	4	0	0	5	1	0	0	0	6	9	1
12	1940_fs_I_guess_ILL_have_to_dream_the_rest.txt.token.f	145	76	1	1	11	0	0	11	8	0	0	0	1	17	9
13	1940_fs_I_havent_the_time_.txt.token.tag	92	62	2	0	10	0	0	9	8	0	0	0	0	19	3
14	1940_fs_I_kissed_your_hand.txt.token.tag	57	37	2	0	3	0	0	5	4	0	0	0	0	11	6
15	1940_fs_I_love_my_baby.txt.token.tag	121	60	6	3	4	0	0	2	7	0	0	0	1	22	0
16	1940_fs_I_married_an_angel.txt.token.tag	67	45	1	0	8	0	0	6	8	0	0	0	1	10	1
17	1940_fs_I_saw_your_face_in_a_cloud.txt.token.tag	131	64	6	1	9	0	0	13	7	0	0	0	0	26	2
18	1940_fs_I_tried.txt.token.tag	81	56	5	2	3	0	0	6	2	0	0	0	0	11	2
19	1940_fs_I_wanna_be_loved_by_you.txt.token.tag	189	37	12	0	2	0	0	14	5	2	0	0	0	44	0
20	1940_fs_Id_know_anywhere.txt.token.tag	64	40	0	0	3	0	0	6	8	0	0	0	3	5	1
21	1940_fs_Im_always_chasing_rainbows.txt.token.tag	120	74	6	0	15	0	0	8	3	0	0	0	2	15	9
22	1940_fs_Im_in_love.txt.token.tag	317	76	4	0	22	6	0	51	15	0	0	0	1	64	3
23	1940_fs_Im_in_the_mood_for_love.txt.token.tag	91	63	3	1	8	0	0	20	4	0	0	0	2	15	3
24	1940_fs_In_a_little_Spanish_town.txt.token.tag	76	50	3	0	11	0	0	11	8	0	0	0	0	11	7
25	1940_fs_Is_there_a_chance_for_me.txt.token.tag	170	50	2	0	21	2	0	20	4	2	0	0	10	37	0
26	1940_fs_Ive_got_a_restless_spell.txt.token.tag	93	55	5	0	5	0	0	6	8	0	0	0	1	8	2
27	1940_fs_Ive_got_my_eyes_on_you.txt.token.tag	51	35	0	0	2	0	0	8	1	0	1	0	0	3	4
28	1940_fs_Ive_lost_my_heart_again.txt.token.tag	164	58	3	0	10	0	0	8	6	0	0	0	1	27	4
29	1940_fs_Just_you_just_me.txt.token.tag	93	55	1	0	3	0	1	8	6	0	0	0	1	10	6
30	1940_fs_Learn_to_croon.txt.token.tag	56	42	2	0	2	0	0	5	4	1	0	0	0	8	2
31	1940_fs_The_sky_fell_down.txt.token.tag	71	55	3	0	8	1	0	5	2	0	0	0	0	10	4
32	1940_fs_Thou_swell.txt.token.tag	72	53	3	2	4	0	0	5	10	0	0	0	1	9	8

figura 2: imagem da planilha do Excel com as 32 primeiras linhas de nossa planilha, contendo as 15 primeiras variáveis (tokens; types; CC; CD ; DT; EX; FW; IN; JJ; JJS; LS; MD; NN; NNS)

Os números em cada coluna representam a quantidade de ocorrência de cada variável em cada arquivo. A primeira linha representa o arquivo intitulado “1940\_fs\_April\_played\_a\_fiddle.txt.token.tag”, que corresponde à música “April played a fiddle”, gravada por Frank Sinatra em 1940. Conforme representado na planilha, a letra dessa música tem 85 ocorrências (*tokens*) de 53 formas (*types*). A planilha também mostra que a música tem 9 ocorrências de conjunções coordenativas (coluna CC); nenhum número cardinal (coluna CD) e 10 ocorrências de determinantes

(coluna DT)<sup>140</sup>.

O mesmo pode ser dito dos arquivos nas linhas seguintes, dadas as devidas diferenças classificatórias e quantitativas. Se a planilha estivesse totalmente representada no quadro anterior, verificaríamos também a quantidade de ocorrência de cada uma das 97 variáveis.

Apresentadas as variáveis e seu processo de seleção, o próximo passo é a análise fatorial. É preciso esclarecer que, para que essa análise seja possível, é necessária a normalização dos dados, ou seja, os valores presentes em cada linha, de cada coluna, devem ser normalizados pela quantidade de ocorrências em cada canção. Isso porque, como dissemos anteriormente, os dados precisam ser comparáveis.

Veja as dez primeiras linhas do quadro anterior com as frequências normalizadas, seguido de explicação mais detalhada:

file	tokens	types	CC	CD	DT
1940_fs_April_played_a_fiddle.txt.token.tag	85	53	105,88	-	117,65
1940_fs_Empty_saddles.txt.token.tag	107	56	9,35	-	93,46
1940_fs_Every_night_about_this_time.txt.token.tag	124	50	24,19	-	169,35
1940_fs_Hands_across_the_table.txt.token.tag	43	35	23,26	-	69,77
1940_fs_Head_on_my_pillow.txt.token.tag	78	50	25,64	-	25,64
1940_fs_ILL_Take_talulah.txt.token.tag	124	73	8,06	8,06	104,84
1940_fs_ILL_never_let_a_day_pass_by.txt.token.tag	90	61	55,56	11,11	111,11
1940_fs_ILL_never_smile_again.txt.token.tag	89	42	-	-	11,24
1940_fs_ILL_string_along_with_you.txt.token.tag	98	56	40,82	20,41	81,63

quadro 18: primeiras linhas de nossa planilha com dados normalizados

O cálculo de normalização é feito dividindo-se o valor da ocorrência, pelo número de palavras em cada texto (valor de *tokens*), multiplicado por mil. Ou seja, no caso da primeira linha (arquivo) em relação à coluna CC, se o arquivo que contém 85 palavras contivesse 1.000, as 9 ocorrências de conjunções coordenativas seriam, na realidade, 105,88. Dessa forma, todas as variáveis de todas as letras de música apresentam valores considerando que o arquivo contivesse 1.000 palavras, conseqüentemente compartilhando o mesmo parâmetro de comparação.

<sup>140</sup> para outros exemplos ver anexo 2, que apresenta as 98 colunas em 30 linhas



Após a normalização das frequências das variáveis, os dados são transferidos para o programa estatístico, para que iniciemos o processo de extração de fatores.

### 4.3. Extração de Fatores

A extração dos fatores é feita por um conjunto de procedimentos estatísticos que, no caso desta pesquisa, foram executados no programa de análise fatorial SPSS versão 17 (Statistical Package for the Social Sciences). O programa processa automaticamente cálculos de estatística descritiva (tabulação cruzada, frequência, descritivas); estatística bivariada (médias, correlações bivariadas, parciais, t-test, etc); previsão numérica (regressão linear); e previsão de identificação de grupos (análise fatorial, análise de clusters). A vantagem do SPSS é que sua interface é bastante amigável ao usuário, ainda que faltem a este conhecimentos profundos da área. É preciso mencionar que Biber (1988) não utiliza o SPSS, mas outro programa, o SAS, para a análise fatorial. Contudo, o resultado é o mesmo.

O SPSS permite a importação da planilha do Excel com os dados da pesquisa, da forma como descrevemos anteriormente e os converte automaticamente para a forma utilizada pelo programa (extensão. sav).

O primeiro passo é definir o método de extração dos fatores. Biber (1988, p. 82) apontou os dois métodos mais comuns: o PFA (*principal factor analysis*<sup>141</sup>), utilizado por Biber, que verifica apenas a variação compartilhada, e o PCA (*principal components analysis*), que considera toda a variação nos dados. Decidimos pelo segundo método de extração, o PCA, levando em consideração a extensão de nossos dados e sua crescente utilização em pesquisas (PACHECO DE OLIVEIRA, 1997; SHERGUE, 2003; CROSSLEY; LOUWERSE, 2007).

O PCA é muito usado em pesquisas exploratórias, isto é, quando o pesquisador não parte de funções determinadas *a priori*, mas utiliza-se do método para reduzir o grande número de variáveis a um número de dimensões subjacentes, como é o caso da nossa pesquisa. Trata-se do método mais usado para extração de fatores porque, segundo Thompson e Vidal-Brown (2001, p. 7), uma característica importante é que ele faz

---

<sup>141</sup> no programa SPSS é chamado de PAF (Principal Axis Factoring)

correlações das médias dos escores entre si, ou seja calcula a média das ocorrências de cada variável em relação a cada uma das outras. Sendo esse o nosso objetivo, definimos o método e passamos para a interação com o programa SPSS. O primeiro passo é seleccionar o botão “analyze” e as opções “dimension reduction” e “factor”, o que retorna a seguinte janela:

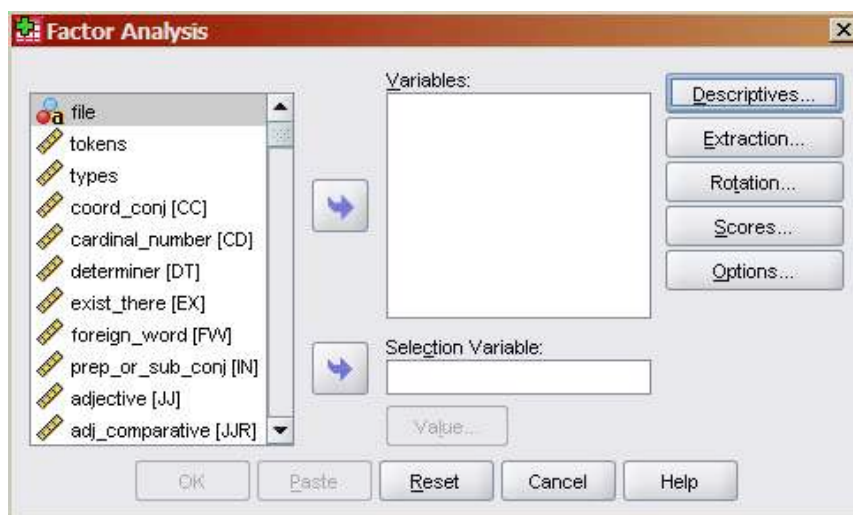


figura 3: imagem da primeira janela do programa de análise fatorial SPSS, para a seleção de variáveis

As variáveis da pesquisa (características morfossintáticas, semânticas e as contagens) estão todas dispostas na primeira coluna, embora não seja possível sua visualização total na figura acima, pois elas dependem da barra de rolagem na janela. As variáveis a serem consideradas para cada análise são transportadas para o quadro “variables”, assim que selecionadas, clicando na setinha superior.

A seleção das variáveis está ilustrada na imagem seguinte:

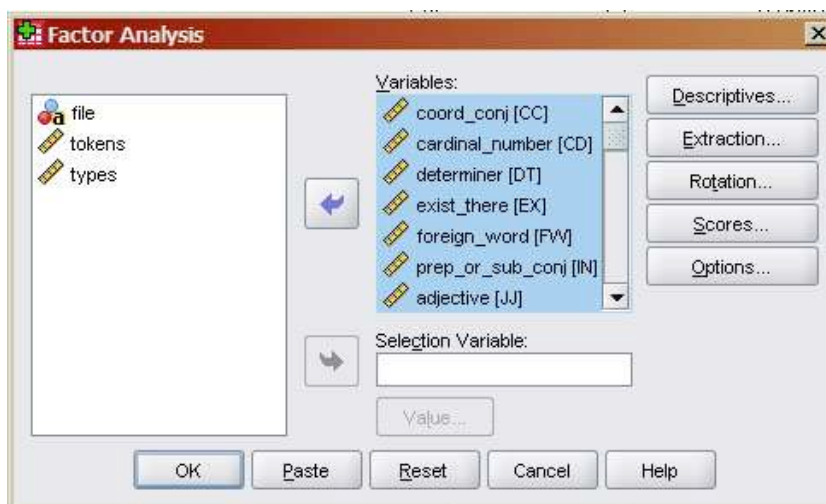


figura 4: Variáveis selecionadas para a análise fatorial

Podemos observar apenas sete das 97 variáveis da pesquisa, mas as outras podem ser vistas quando movimentada para baixo a barra de rolagem ao lado da coluna “variables”. As variáveis *types* e *tokens* também foram incluídas em seguida. O próximo passo é definir os parâmetros de análise apresentados pelas cinco opções (botões) à direita: “*descriptives, extraction, rotation, scores, options*”. Cada uma dessas opções leva a uma outra janela, com as possibilidades de análise em cada nível.

Seguimos os passos de Biber (1988) e Kauffmann (2005) selecionando para “*descriptives*”, ou análise descritiva, os seguintes itens : “*initial solution*” e a matriz de correlação “*KMO and Bartlett’s test of sphericity*” e “*anti-image*”. Essas seleções são feitas a fim de se verificar a adequação dos dados em termos de correlação. O KMO quantifica as relações entre as variáveis a partir da totalidade dos dados e o teste de Bartlett verifica a possibilidade da não correlação entre as variáveis. O teste de Bartlett é um teste geral que verifica se a análise não é inválida por existir correlação entre as variáveis.

O próximo botão é o “*extraction*”, onde apontamos o método de extração (PCA), de análise (*correlation matrix*), a forma de *display*, a extração e o número de corte para convergências. A tabela *default* do programa está disposta a seguir:

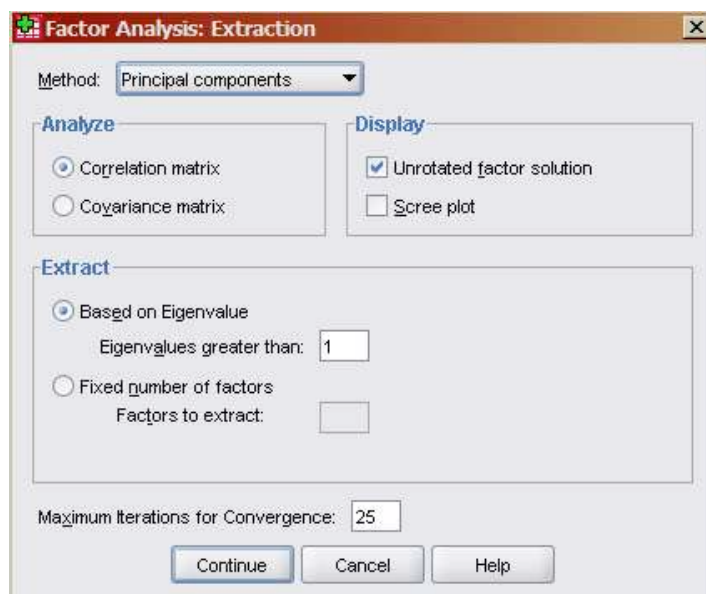


figura 5: imagem da janela de extração de análise fatorial, com suas seleções padrão (*default*)

Selecionamos as opções padrão (*default*) para quase todas as possibilidades, menos o *display*, pois precisávamos observar a distribuição de fatores no gráfico *scree*. Selecionamos uma matriz de correlação e não uma matriz de covariância, pois o nosso objetivo é verificar como as variáveis se relacionam entre si. O gráfico *scree*, frequentemente usado em análises fatoriais do tipo PCA (KAUFFMANN, 2005, p. 72), é um indicativo auxiliar na determinação do número de fatores, pois a curva produzida pelo gráfico vai indicar o número de fatores, a partir da observação dos pontos onde a curva deixa de ser pronunciada. Nesses pontos os valores *Eigen*<sup>142</sup>, selecionados para a extração, deixam de ser relevantes, pois apresentam pouca diferença. Esse gráfico também é acompanhado por uma tabela chamada “*total variance explained*” e, juntos, apresentam os fatores que irão explicar a variação dos dados no corpus (na coluna “*Initial Eigenvalues*”) – calculados a partir da soma das variações parciais cumulativamente até 100% do total de variação.

O resultado desse cálculo indica o percentual de variação acumulada a partir de uma determinada quantidade de fatores. Portanto, é necessário que o cálculo seja feito novamente selecionando-se o número de fatores que se quer obter. Esse procedimento faz com que os cálculos sejam feitos diversas vezes, até que se obtenha a quantidade de

<sup>142</sup> “Eigenvalue é o índice que representa a quantidade de variância compartilhada por cada fator (Lee, 2000, p.37)” (KAUFFMANN, 2005, p. 21)

fatores que apresentam alto percentual de variação acumulada. No caso de nossa pesquisa, a extração se deu em três etapas, que serão explicadas na seção seguinte.

O próximo passo é selecionar o método de rotação, para o qual escolhemos o Promax (Kappa 4), o mesmo selecionado por Biber (1988). A rotação é necessária para que os dados sejam mais claramente interpretados, pois a matriz fatorial originalmente é muito complicada para a interpretação, visto que os fatores, ao serem rotacionados, tornam-se o mais distinto possível dos demais. Ou seja, na rotação, as variáveis de cada fator são dispostas de tal modo a realçar as diferenças entre cada fator. Em geral, cada fator representa a maior proporção de variação possível, sobrecarregando um deles com a maior quantidade de variáveis. Para evitar-se que isso ocorra, é necessário um processo de rotação oblíqua que equilibrará essa carga. O método Promax permite uma menor correlação entre os fatores enquanto aumenta a intercorrelação entre eles. Biber recomenda o Promax para a descrição de variação textual”, pois quando “os fatores representam dimensões textuais subjacentes, não há razão para se assumir que os fatores estejam completamente não relacionados” (BIBER, 1988, p. 85), que foi, conforme dito, o escolhido para esta pesquisa.

O próximo botão (opção) “scores” não é selecionado na primeira etapa da extração, pois nossa extração de escores será feita posteriormente à obtenção dos fatores, isto é, quando tivermos definido o número de fatores para a nossa análise, extrairemos seus valores.

Para o último botão, ‘options’, na área “missing values” selecionamos “exclude cases listwise” e na área “coefficient display format” selecionamos “sorted by size”, com valores absolutos inferiores a 0,30, assim como Biber (1988), pois, quanto mais próximo de 0 menos interação existe e, conseqüentemente, não interessa a esta pesquisa. A seleção por tamanho facilita a leitura da tabela de resultados.

O SPSS processou os dados muito rapidamente e retornou uma janela de “output” contendo os seguintes dados:

a) *Title: Factor Analysis*

b) *Notes:* apresenta os dados da análise, como o nome do arquivo, as variáveis selecionadas, o método escolhido, etc.

- c) *Active database*: nome do arquivo. Em nosso caso “lmda\_normal.sav”
- d) *KMO and Bartlett’s Test*: tabela com os resultados desses testes.
- e) *Anti-image Matrices*: duas matrizes de correlação entre cada variável. Interessamos apenas a segunda, que apresenta o resultado anti-imagem, conferindo um valor a cada uma das variáveis em relação às outras.
- f) *Communalities*: uma tabela com o índice de comunalidade inicial das variáveis, ou seja, o índice de correlacionamento de uma variável com as demais
- g) *Total variance explained*: uma tabela com valores para todas as variáveis, que explica a variação que justifica o agrupamento das variáveis em fatores. O maior índice de variância fica no primeiro fator, composto pelas variáveis mais fortes. Conforme o índice vai diminuindo, são computados os fatores seguintes, até que toda a variância seja explicada (agrupada) por fatores. Essa variância também é demonstrada graficamente no item seguinte.
- h) *scree plot*: demonstra graficamente os valores obtidos (“*eigenvalue*”), distribuídos em fatores. A observação da curva produzida no gráfico pelos valores ‘*Eingen*’ determina o número de fatores. Os pontos onde o gráfico apresenta mudança de variação indicam possíveis fatores, os quais aparecem especificadamente no próximo item.

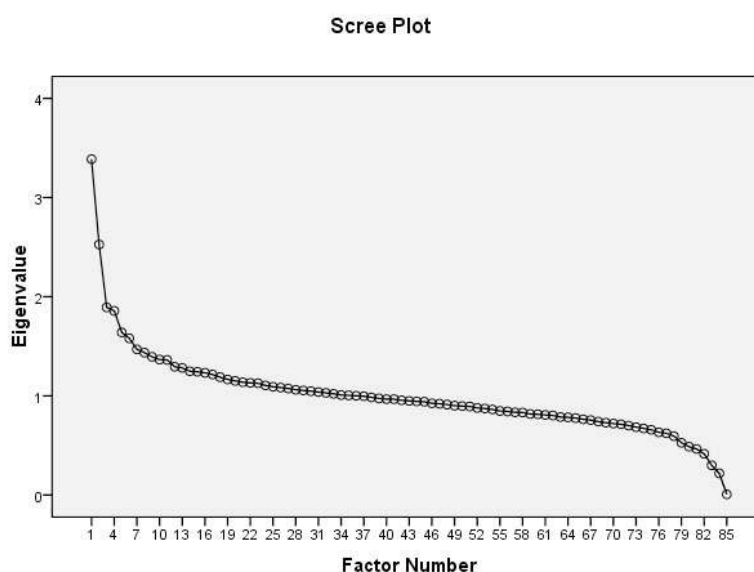


figura 6: Imagem de um gráfico *scree* com todas as variáveis de nossa pesquisa

h) *Component Matrix*: indicação do número de fatores.

i) *Pattern Matrix*: Essa é a matriz mais importante, pois ela apresenta os valores de cada variável para cada componente (fator). Seus dados são usados para a extração dos escores dos fatores que servirão de parâmetro para os arquivos do corpus. Os valores simbolizam os pesos das variáveis em cada fator. Devem ser consideradas, para a interpretação, apenas o maior valor. Os valores também aparecem acompanhados pelos sinais de negativo e positivo, os quais indicam oposição de frequência. Isto é, a variável com um dos sinais (positivo ou negativo) encontra-se em posição oposta à outra reapresentada pelo sinal inverso. Em outras palavras, quando uma é bastante frequente a outra é pouco. (Voltaremos a falar sobre isso quando apresentarmos os dados)

j) *Structural Matrix*: embora não seja usada para esta análise, a extração de fatores também apresenta uma matriz estrutural que reflete mais complexamente a matriz padrão.

i) *Component Correlation Matrix*: O último dado da extração fatorial é a matriz de correlação de componentes, que demonstra a proporção de correlação entre os fatores

A explicação de cada componente da análise fatorial contido no *output* do SPSS fica além de nosso propósito, pois nos munimos do programa como usuários e não como especialistas em estatística. Embora seja útil saber a fundo cada uma das opções do SPSS, o programa é tão complexo que apenas *experts* em estatística possuem tal conhecimento. Com efeito, o programa é oferecido como uma ferramenta, cujo objetivo é permitir que usuários de diversas áreas possam efetuar análises estatísticas sem precisar conhecer minuciosamente a matemática envolvida no processamento dos dados.

Até o momento, descrevemos os passos que tomamos sem apresentar os resultados obtidos a partir de nossos dados, pois julgamos pertinente essa breve explanação sobre os procedimentos estatísticos. A seguir, apresentamos nossa extração fatorial, que ainda constitui parte de nossos procedimentos metodológicos, pois, como já mencionamos, muitas de nossas decisões metodológicas foram tomadas a partir de resultados obtidos

dos procedimentos anteriores. Ou seja, as dimensões de análise de letras de música anglo-americana que obtivemos como resultado final de nossa análise são fruto da interpretação de fatores produzidos pela análise fatorial, a qual envolve uma série de procedimentos e decisões metodológicas interdependentes. Em nosso caso, a extração fatorial se deu em duas etapas. A primeira etapa indicou a necessidade do agrupamento de variáveis, a segunda indicou o número de fatores e a extração em três passos diferentes, que estão explicitados na seção seguinte.

#### 4.3.1. Primeira Extração

Nossa primeira tentativa de extração de fatores mostrou-se bastante complexa. Os dados foram inseridos no programa SPSS da maneira como descrevemos na seção anterior, assim como os passos e determinações para a análise estatística. O gráfico “*scree plot*” resultante é o mesmo disposto na última figura apresentada na seção anterior (figura 6), no qual percebemos uma forte sobreposição de variáveis e que foi resumido pela seguinte Matriz Fatorial:

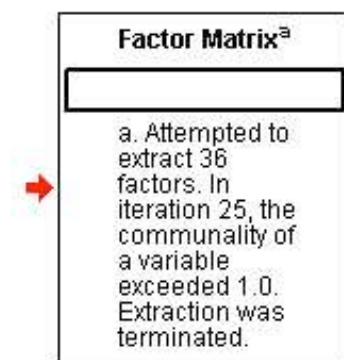


figura 7: Imagem da matrix fatorial resultante de nossa primeira extração

O programa tentou extrair 36 fatores, em 25 interações. O índice de comunalidade<sup>143</sup> de cada variável excedeu 1,0, o que significa que existe possibilidade

<sup>143</sup> índice que indica, quando “inicial” o nível de relacionamento de cada uma das variáveis com a outra, e quando “final” a proporção de variância explicada pelos faores extraídos. “Uma variável com baixa comunalidade (menor que 0,10) indica ao pesquisador que “preciso considerar a possibilidade de sua exclusão da Análise Fatorial” (KAUFFMANN, 2005, p. 71)



de correlação entre as variáveis. Porém, 36 fatores é um número muito grande e a observação do gráfico *scree* sugere que a distribuição das variáveis ocorre de forma muita aglomerada entre os fatores, o que dificulta uma análise precisa.

Essa primeira análise dos dados forneceu indicações de que poderíamos ter uma visão mais específica do comportamento de características linguísticas no corpus, se considerássemos o agrupamento de variáveis, especialmente quando levamos em consideração a natureza de nossos dados (pouca quantidade de palavras por texto) e sua extensão (mais de 6000 letras de música). Dessa forma, reorganizamos as 97 variáveis iniciais nas seguintes variáveis de análise expostas a seguir:

1. *song bundles*
2. *google count*
3. *noun* – todos os tipos de substantivos
4. *person-pron* – pronomes pessoais
5. *possessives* – todas as indicações de posse em inglês: pronomes e adjetivos possessivos, ‘s e ‘whose’
6. *past tense* – verbos em todas as formas de passado
7. *present tense* – verbos em todas as formas de presente
8. *infinitive-gerund* – verbos no infinitivo e no imperativo
9. *qualifiers* – todos os determinantes, adjetivos e comparativos
10. *prepositions* – todas as preposições
11. *coord-conj* – conjunções e palavras coordenativas (wh)
12. *interrog* – marcadores interrogativos
13. *subord* – marcadores de subordinação
14. *modals* - modalizadores
15. *interj* – interjeições
16. *movement* – palavras que envolvem movimento contidas nas categorias: *bodymotion, motion, process, radiant light, impacting, intentional process, walking e pursuing*
17. *emotion* – *capability, emotional state, intent-psychological process, needs, proposition, true, unhappiness, wants, fantastic e religion*
18. *musical manifestation* – tudo que envolve música e ritmo: *dancing, music, singing, radiant sound*
19. *time* – manifestações de tempo
20. *speech* – *requesting, speaking, stating*
21. *people* – *human, male, female, group of people, social role*
22. *object* – todo os tipos de objeto
23. *social* – problemas sociais – *drugs, sex, violence*

Os agrupamentos foram feitos de modo que se pudesse perceber gramaticalmente

quais músicas falam mais sobre o passado, ou sobre o presente, quais se concentram mais em pessoas, ou objetos, quais apresentam mais expressões de interjeição e assim por diante. Três categorias foram desprezadas por terem baixíssima representatividade no corpus: listas, símbolos e palavras estrangeiras. O número de palavras diferentes (*types*) e de ocorrências (*tokens*) também não foi considerado para a análise fatorial, embora essas duas características sejam eventualmente retomadas para a análise e, a fim de que pudessem ser observadas com maior facilidade, foram mantidas nas tabelas de variáveis enviadas ao programa SPSS.

Semanticamente, os agrupamentos foram feitos por meio da observação manual de cada uma das palavras que faziam parte de cada categoria e agrupadas por afinidade. Por exemplo, todos os hipônimos sob o hiperônimo *emotional state* (tais como “fear, shame, lonely, thrill, sorry, happiness”, etc) foram observados em relação aos outros das categorias *capability* (i.e., *care, sigh, fantasy, imagination, etc*); *intent-psychological process* (i.e., *dream, remember, forget, worry, etc.*); *needs* (*need, want, involve*); *proposition* (*faith, thought, sense, idea, etc*); *true* (*really, true, real, truth, etc.*); *unhappiness* (*pain, sad, sorrow, heartache, etc.*); *wants* (*love, desire, want, wish, etc.*); *fantastic* (*mith, mystical, sandman, Alladin, etc*); e *religion* (*amaze, arise, baptism, christian, etc.*). Vale ressaltar que algumas dessas categorias possuem mais de 100 palavras. O Anexo 4 apresenta um exemplo de uma categoria semântica (*emotional state*) com todas as palavras.

Como já dissemos, as características analisadas desempenham funções textuais, que serão observadas em nossa análise dos fatores e das dimensões de letras de música. As variáveis semânticas indicam os temas tratados nas letras de música já indicados pelo próprio nome da variável. Já no caso das variáveis gramaticais, as funções desempenhadas podem ser melhor visualizadas no quadro a seguir:

Variável	Funções	Fonte
<i>song_bundles</i>	Indicam a proximidade da linguagem das letras de música com a linguagem geral, em termos de sequências de três palavras.	Biber; Conrad; Cortes, 2004

Variável	Funções	Fonte
<i>google counts</i> (variável original desta pesquisa)	Indicam a proximidade da linguagem das letras de música com a linguagem geral, em termos de sequências de três palavras	Crossley; Louwerse, 2007
substantivos	Carregam informação, algumas nominalizações são usadas para reduzir informações abstratas. No caso de substantivos próprios indicam referências extra textuais.	Biber (1988, p. 227-228); Pacheco de Oliveira (1997, p.92)
pronomes pessoais	Marcam o envolvimento interpessoal. <sup>144</sup>	Pacheco de Oliveira (1997, p.93); Biber (1988, p.225)
possessivos	Marcas de posse dos referentes envolvidos.	Biber (1988, p.225)
verbos no passado	Marcadores de produção <b>narrativa</b> e narrativa/descritiva no <b>passado</b> , com foco na sequência temporal.	Biber (1988, p.223)
verbos no presente	Marcam tópicos e ações de relevância <b>imediate</b> . Seu uso está associado a discursos não planejados.	Biber (1988, p. 224)
verbos no infinitivo, gerúndio e imperativo	Infinitivos são vistos como típicos da forma escrita, usados para integrar e unificar o texto, quando associados a outros verbos. Gerúndios desempenham função de marcadores de tempo imediato. Imperativos são usados para se dirigir ao leitor através do sujeito de segunda pessoa.	Biber (1988, p. 227; 232); Pacheco de Oliveira (1997, p. 93)
qualificadores	Marcam elos e referências, sejam emocionais ou atributivas. São usados para expandir e elaborar informações presentes no texto.	Pacheco de Oliveira (1997 p. 94; 96) Biber (1988, p. 237) Kauffmann (2005, p. 84)
preposições	Usadas para ligar informações, integrar idéias, tendem a ocorrer em discursos do tipo informativo (cartas, acadêmico, documentos oficiais, etc.).	Biber (1988, p.237; 1995)
coordenação	Indicam integração entre frases; (marcam diferenças e criar fragmentações)	Biber (1988, p. 245) Pacheco de Oliveira (1997, p. 95); Kauffmann (2005, p. 84)

<sup>144</sup> Primeiras e segundas pessoas participantes da comunicação; terceiras pessoas marcam referências externas

Variável	Funções	Fonte
interrogações	Marcam perguntas; indicando preocupação com funções interpessoais.	Biber (1998, p. 227); Pacheco de Oliveira (1997, p. 93)
subordinação	Indicam conexão entre orações e complexidade estrutural.	Pacheco de Oliveira (1997, p. 95); Biber, (1988, p. 229)
modais	marcam permissão, possibilidade, habilidade, obrigação, futuro e predição.	Biber (1988, p. 241; 2006, p.92) Pacheco de Oliveira (1997, p. 94)
interjeições	demonstram reações afetivas dos falantes;	O'Keefe; MacCarthy; Carter (2007, p. 172)

quadro 19: funções desempenhadas pelas variáveis gramaticais

Considerando as funções das variáveis semânticas e gramaticais, após a extração fatorial emergirão dimensões de variação linguística nas letras de música. Uma vez definidas e agrupadas as variáveis, nossa tabela inicial no programa Excel, que continha 6.291 linhas e 98 colunas, teve seu número de colunas reduzido para 26 colunas. A imagem abaixo apresenta os 10 primeiros arquivos com as frequências normalizadas para as 10 primeiras colunas. O Anexo 3 apresenta as 30 primeiras linhas do arquivo com todas as 26 colunas (25 variáveis mais o nome dos arquivos).

	A	B	C	D	E	F	G	H	I	J	K
1	file	tokens	types	songbundles	google_coun	past_tense	present_tense	infinitive_gerund	interj	qualifiers	possessives
2	1940_fs_April_played_a_fiddle.txt.token.tag	85	53	11,76	258,82	141,18	23,53	47,06	0,00	188,24	117,65
3	1940_fs_Empty_saddles.txt.token.tag	107	56	0,00	93,46	28,04	130,84	65,42	0,00	261,68	37,38
4	1940_fs_Every_night_about_this_time.txt.token.tag	124	50	32,26	201,61	0,00	96,77	56,45	2,42	241,94	137,10
5	1940_fs_Hands_across_the_table.txt.token.tag	43	35	69,77	325,58	0,00	93,02	93,02	0,00	186,05	209,30
6	1940_fs_Head_on_my_pillow.txt.token.tag	78	50	12,82	217,95	25,64	102,56	64,10	0,00	102,56	243,59
7	1940_fs_ILL_Take_talulah.txt.token.tag	124	73	8,06	177,42	56,45	120,97	80,85	0,00	145,16	104,84
8	1940_fs_ILL_never_let_a_day_pass_by.txt.token.tag	90	61	22,22	177,78	44,44	22,22	111,11	0,00	233,33	111,11
9	1940_fs_ILL_never_smile_again.txt.token.tag	89	42	78,85	134,83	0,00	112,36	67,42	0,00	202,25	213,48
10	1940_fs_ILL_string_along_with_you.txt.token.tag	98	56	10,20	295,92	10,20	71,43	71,43	0,00	255,10	122,45
11	1940_fs_I_could_make_you_care.txt.token.tag	78	46	38,46	256,41	102,56	25,64	166,67	0,00	115,38	294,87

figura 8: Amostra da planilha para análise fatorial com variáveis agrupadas e dados normalizados

As nossas categorias apresentam-se agrupadas, portanto em três diferentes tipos: gramaticalmente, semanticamente e por pacotes lexicais (*bundles*). Na categoria gramatical encontram-se as variáveis relativas a substantivos; possessivos; infinitivos e gerúndios; modais; tempo presente; tempo passado; subordinações; qualificadores;

coordenações e conjunções; pronomes pessoais; preposições; interjeições e interrogações. Na categoria semântica, encontram-se agrupadas palavras que indicam movimento; emoções; pessoas; tempo; discurso/fala; objeto; manifestação musical; e referências sociais. Na categoria de *bundles* encontram-se as contagens dos trigramas presentes nas letras de música (*song bundles*); e da frequência dos trigramas das letras de música no corpus do Google N-Grams (*google counts*). Considerando essa divisão, optamos por analisar os fatores sob a mesma perspectiva, ou seja, analisar as variáveis em três instâncias: gramatical (apenas com as variáveis gramaticais), semântica (apenas com as variáveis semânticas) e léxico-gramatical (associando a contagem de frequência de trigramas às variáveis gramaticais).

#### **4.3.2. Extração em três partes**

O processo de extração fatorial seguiu exatamente os mesmos passos descritos anteriormente (item 4.3). A diferença esteve na seleção de variáveis. Se naquela selecionamos todas as variáveis de uma única vez, nesta selecionamos as variáveis agrupadas e divididas em três categorias. Os procedimentos de extração estão descritos a seguir.

##### **4.3.2.1. Componentes gramaticais**

A extração de componentes gramaticais no SPSS seguiu os mesmos passos, porém, na janela de seleção de variáveis, escolhemos apenas as 13 que integraram esta análise de cunho gramatical.

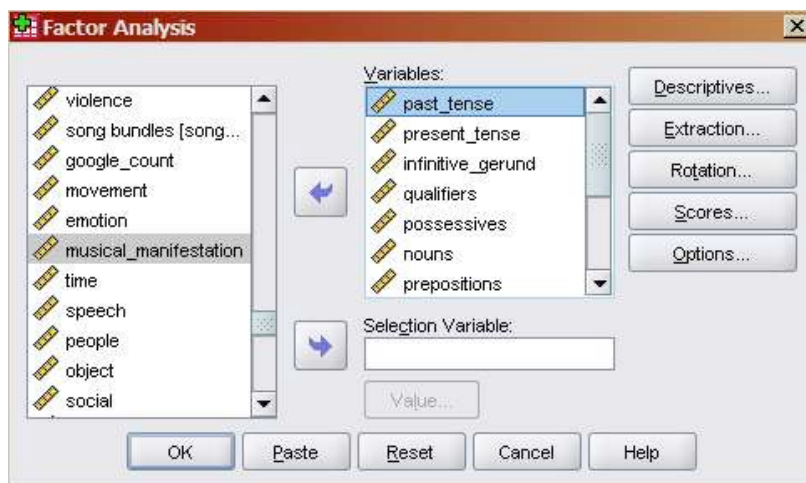


figura 9: Quadro de análise fatorial com variáveis gramaticais selecionadas

O quadro mostra 7 das 13 variáveis de análise. Para os botões “*descriptives*”, “*extraction*”, “*rotation*”, “*scores*” e “*options*” foram selecionados os mesmos parâmetros apresentados anteriormente. O *output* do SPSS para essa nova extração nos retornou o seguinte gráfico *scree*:

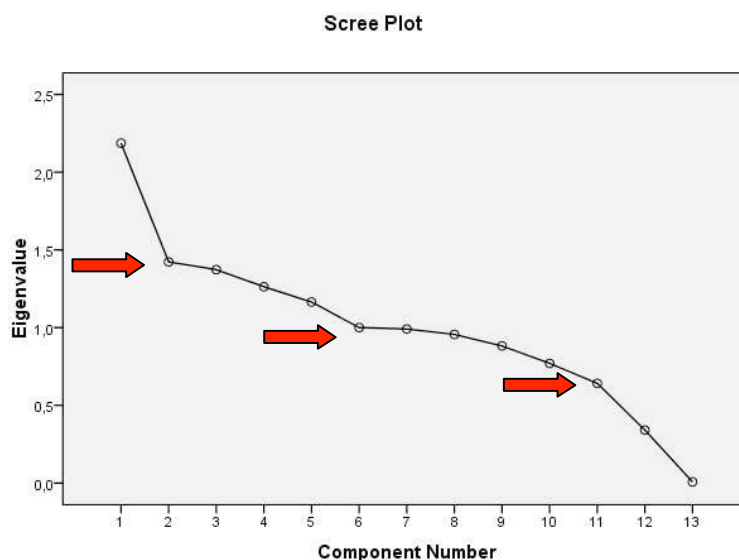


figura 10: Gráfico *scree* produto da extração dos componentes gramaticais

Nosso gráfico demonstra três pontos onde a curva deixa de ficar mais pronunciada, que podem ser vistos nas seguintes posições: entre os componentes 1 e 2, na direção do

componente 6 e na direção do componente 11. Essas marcas sugerem três fatores. A partir dessa constatação, foi feita nova extração considerando-se três fatores, o que resultou na matriz padrão reproduzida a seguir, contendo o valor de concorrência média (escore) para cada variável analisada em cada fator, considerando o ponto de corte de 0,30 por nós selecionado (cf. Biber, 1988, p. 103 e item 4.3 deste capítulo). O escore de fatores é calculado a partir da somatória das ocorrências das características que apresentam carga salientes em cada fator.

**Pattern Matrix<sup>a</sup>**

	Component		
	1	2	3
nouns	-,749		
infinitive_gerund	,636		
modals	,615		
prepositions	-,306		
subord			
personal_pron	-,401	,775	
qualifiers		-,729	
possessives	,364	,719	
past_tense			,819
present_tense			-,747
coord_conj			,316
interrog			

Extraction Method: Principal Component Analysis.  
Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 6 iterations.

figura 11: Matriz padrão da extração de componentes gramaticais com 13 variáveis, 3 fatores (pesos acima de 0,30)

Das treze variáveis consideradas, duas não apresentam pesos (valores) na imagem anterior, porque se encontram abaixo do ponto de corte em todos os fatores: as conjunções subordinativas e as interrogativas, ou seja o valor calculado para elas foi inferior ao estipulado como limite para valores absolutos (0,30) na janela de opções da análise fatorial.

Biber (1988, p.87) nos lembra que existem diferentes técnicas para a observação dos pesos (*loadings*) de uma matriz padrão para análise fatorial, porém “pesos que possuem um valor absoluto inferior a 0,30 são geralmente excluídos como não importantes, ainda

que sejam estatisticamente significantes”<sup>145</sup>, conforme mencionamos no início da seção 4.3.

Seguindo ainda as recomendações de Biber (1988, p. 93), embora tenhamos estipulado 0,30 como ponto de corte, consideraremos para a análise apenas aquelas que apresentam peso superior a 0,35. Dessa forma, estamos privilegiando as variáveis que representam os fatores com maior saliência. Por conseguinte, as variáveis preposição e conjunções coordenativas também não foram consideradas para análise, pois seus valores são (-,306) e (0,316), respectivamente.

O sinal positivo ou negativo não interfere no grau de importância do valor. Os sinais não demonstram inferioridade ou superioridade, mas agrupam as características mais (+) e menos (-) frequentes quando aparecem juntas.

Outro ponto importante é que quando uma característica ocorre em componentes diferentes, deve ser considerada apenas naquele em que o valor é maior. Esse é o caso de pronomes pessoais e possessivos que possuem valores nos componentes 1 e 2, mas, em ambos os casos, os valores no componente 2 são superiores ao componente 1, ou seja, (0,401 X 0,775) e (0,364 X 0,719), respectivamente. Lembrando que, neste momento, o sinal de negativo que acompanha o valor (-0,401) não deve ser ainda considerado.

A matriz padrão mostra três variáveis para o componente 1: substantivos; infinitivos/imperativos e modais. Os índices de polaridade, positivo e negativo (sinais de + e -), indicam que quando infinitivos/imperativos e modais são bastante presentes numa letra de música, a ocorrência de substantivos é baixa. Ao mesmo tempo, quando uma letra tem muitos substantivos ela terá baixa ocorrência de imperativos e modais.

O componente 2 também é formado por três variáveis: pronomes pessoais, qualificadores e possessivos. Os índices de polaridade indicam que uma letra de música representante deste componente, quando apresenta grande ocorrência de pronomes pessoais e possessivos, apresenta também baixa ocorrência de qualificadores e vice e versa.

---

<sup>145</sup> Do original: “... but loadings having an absolute value less than .30 are generally excluded as unimportant even if they are statistically significant”



O componente 3 conterá textos (letras de música) que tipicamente apresentam forte ocorrência de passado, ao mesmo tempo que baixa ocorrência de presente e vice e versa.

Os valores das variáveis presentes nos componentes da análise fatorial fornecem base para a atribuição de escores das variáveis sobre as letras de música presentes em nosso corpus, que leva a um escore por texto/letra de música. Dessa forma, podemos ter um perfil estatístico de cada texto e definir quais textos são tipicamente mais representativos de cada fator.

Esse procedimento será demonstrado no próximo capítulo, após apresentarmos os componentes das nossas três esferas de análise.

#### 4.3.2.2. Componentes semânticos

Extraímos os componentes semânticos seguindo os mesmos passos de extração para os componentes gramaticais, selecionando, contudo, apenas as variáveis de cunho semântico. Essas variáveis compreendem as palavras que indicam movimento, emoção, manifestação musical, tempo, fala, pessoas, objeto e questões sociais.

O resultado da análise do SPSS para essas variáveis retornou um gráfico scree que indicou a existência de três fatores também, como podemos observar no próprio gráfico:

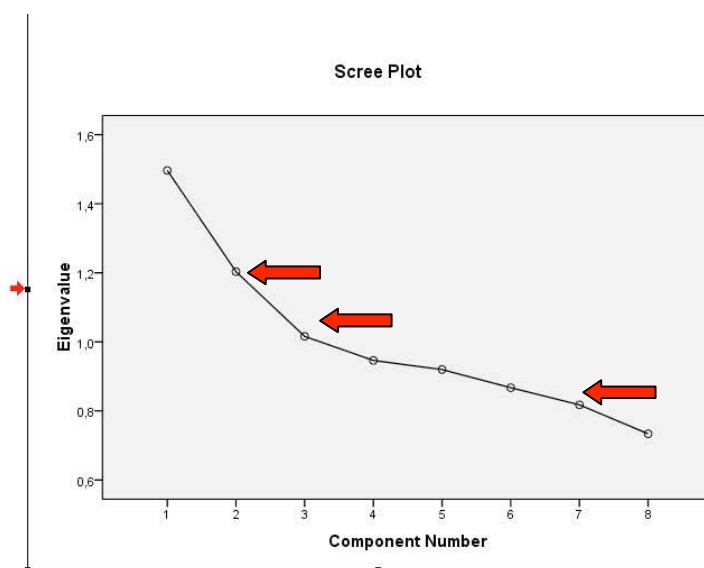


figura 12: Gráfico scree produto da extração dos componentes semânticos

Três pontos marcam os locais onde a curva altera sua intensidade: na altura dos componentes 2, 3 e 7, sugerindo, da mesma forma que para os padrões gramaticais, a existência de três fatores. A matriz padrão resultante da extração para três fatores das variáveis semânticas indica, como veremos a seguir, que todas as variáveis estão acima do valor de corte (0,30) e do valor mínimo absoluto estipulado para a análise (0,35). Sendo assim, diferentemente do que ocorreu com as variáveis gramaticais (as que estiveram abaixo deste último valor não foram consideradas), todas serão consideradas para a análise dos fatores

	Component		
	1	2	3
movement	,632		
emotion		,759	
musical_manifestation			,865
time	,454		-,366
speech	,532		
people	,534		
object	,494		
social		,769	

Extraction Method: Principal Component Analysis.  
Rotation Method: Promax with Kaiser Normalization.  
a. Rotation converged in 4 iterations.

figura 13: Matriz padrão da extração de componentes semânticos com 08 variáveis, 3 fatores (pesos acima de 0,30)

Lembramos mais uma vez que as variáveis só devem ser consideradas para os componentes em que receberam o maior valor, independente do índice de polaridade (positivo ou negativo). A polaridade, como dissemos, indica a oposição de frequência, ou seja, quando em um texto as variáveis do polo positivo são frequentes, as variáveis do polo negativo aparecem com baixa frequência.

Interessantemente, os componentes lexicais não apresentam oposição de polaridade. A única variável com polaridade negativa é a de tempo para o componente 3 (-0,366). Porém, essa mesma variável possui valor maior para o componente 1 (0,454) e deve,

portanto, ser desprezada na computação dos valores para o componente 3.

A análise fatorial para os componentes lexicais em nosso corpus está representada por:

Componente 1: forte presença de palavras que indicam movimento, tempo, fala, pessoas e objetos, aparecendo concomitantemente.

Componente 2: forte presença de palavras associadas a emoção e questões sociais.

Componente 3: presença marcante de palavras associadas a música, dança, canto e som.

O passo posterior à análise fatorial é a extração dos escores de cada um dos componentes (gramaticais, semânticos e léxico-gramaticais) por letra de música, para que se possa verificar quais músicas são mais representativas de cada fator, o que explicaremos adiante. Antes disso, continuaremos nossos procedimentos para extração de fatores com léxico-gramaticais.

#### **4.3.2.3. Componentes léxico-gramaticais**

Quando selecionávamos as variáveis que fariam parte de nossa pesquisa, julgamos que considerar os trigramas, ou pacotes lexicais (*lexical bundles*) de três palavras, mais frequentes em nosso corpus de estudo em relação à suas frequências por texto e em relação ao constaste de suas frequências no *google*, serviriam de fonte indicadora da padronização das letras de música. Assim, consideramos essas frequências como variáveis associadas às categorias gramaticais que, após computadas pelo programa de análise fatorial do SPSS, retornou a seguinte matriz padrão:

Pattern Matrix <sup>a</sup>			
	Component		
	1	2	3
nouns	-.792		
possessives	.460	.649	
qualifiers		-.685	
past_tense			.816
present_tense			-.706
interrog			
modals	.442		
personal_pron	-.373	.813	
infinitive_gerund	.464		
prepositions			
coord_conj			.335
subord			
song bundles	.482		
google_count	.722		

Extraction Method: Principal Component Analysis.  
Rotation Method: Promax with Kaiser Normalization.  
a. Rotation converged in 5 iterations.

figura 14: Matriz padrão da extração de componentes léxico-gramaticais (12 componentes gramaticais mais as contagens de trigramas)

A matriz apresenta três componentes, assim como as matrizes para as variáveis de cunho puramente gramatical e as de cunho lexical. Da mesma forma como nas anteriores, foram desprezados os valores inferiores a 0,35 por representarem menor índice de interação entre as variáveis (BIBER, 1988).

Além disso, conforme visto anteriormente, quando as variáveis têm valores em diferentes componentes, é considerado apenas o de maior valor, independente de sua polaridade negativa ou positiva.

Assim, o componente 1 é representado pela presença das seguintes variáveis: modais, infinitivos/imperativos, *song bundles* (trigramas mais frequentes no corpus) e *google counts*; em oposição à presença de substantivos. Ou seja, um texto representante deste fator quando apresentar forte presença de substantivos terá pouca presença das outras variáveis do componente 1 e vice e versa. Já o componente 2 é representado pela presença de qualificadores em oposição à presença de pronomes pessoais e possessivos.

E o componente 3 apresenta a oposição entre presente e passado.

É importante ressaltar que apesar de incluídas as variáveis de frequência lexical pelo agrupamento de três palavras (trigramas), a formação dos componentes desta esfera de análise não apresenta grande distinção da esfera puramente gramatical. O quadro abaixo possibilita uma visão mais clara da distribuição das variáveis nas duas instâncias:

variável	Componentes gramaticais			componentes léxico-gramaticais		
	1	2	3	1	2	3
nouns	-,749			-,792		
possessives	,364	,719		,460	,649	
qualifiers		-,729			-,685	
past_tense			,819			,816
present_tense			-,747			-,706
interrog						
modals	,615			,442		
personal_pron	-,401	,775		-,373	,813	
infinitive_gerund	,636			,464		
prepositions						
coord_conj						
subord						
song bundles	não incluídas			,482		
google_counts	não incluídas			,722		

quadro 20: quadro comparativo dos pesos das variáveis nos componentes gramaticais e nos componentes léxico-gramaticais

Como podemos observar, nos dois casos, as variáveis com valores inferiores a 0,35, conseqüentemente, ausentes na tabela, são as mesmas: interrogativas; conjunções coordenativas e subordinativas e preposições.

Os valores das outras variáveis, embora diferentes, são comparáveis e muito próximos. A diferença, de fato, deve estar representada no componente 1 da análise de componentes léxico-gramaticais pela presença das duas novas variáveis, as quais se enquadraram simultaneamente no mesmo componente.

Não fosse essa diferença, praticamente, poderíamos ter desprezado a análise neste nível. Contudo, como procuramos também investigar a presença dos trigramas nas músicas, mantivemos as três esferas de análise.

#### 4.4. Fatores

A análise estatística de ordem fatorial de nossos dados apontou três componentes para cada uma das esferas de análise por nós determinadas ao longo da pesquisa, ou seja, três componentes de ordem gramatical (conforme etiquetagem morfosintática do corpus), três componentes de ordem semântica (conforme etiquetagem do corpus por família de palavras) e três componentes de ordem léxico-gramatical (etiquetas gramaticais mais trigramas).

O primeiro passo é a extração dos escores padronizados para cada componente fatorial. Segundo Biber (1988, p. 93-96), esse valor é obtido “somando-se, para cada texto, o número de ocorrências das características com cargas salientes em cada fator”<sup>146</sup>, com base em seus desvios padrões. Esse cálculo é feito automaticamente pelo programa SPSS e pode ser representado pela seguinte expressão matemática:

$$\text{escore padronizado do fator} = \frac{(\text{frequência da variável no texto} - \text{media da variável no corpus})}{\text{desvio padrão da variável}}$$

No SPSS é necessário clicar na opção “*analyze*”, seguida por “*descriptive statistics*”, selecionando a opção “*descriptives*” com “*save standardized scores*” marcado. Esse procedimento gera no arquivo inicial uma nova coluna para cada variável de análise. Para que não se confundam com as originais, seus títulos têm acrescido a letra Z maiúscula no início. A título de ilustração, apresentamos a seguir, um quadro com as 5 primeiras linhas do nosso corpus e três colunas com variáveis com escores padronizados<sup>147</sup>:

file	Zpossessives	Zpast_tense	Zpresent_tense
1940_fs_April_played_a_fiddle.txt.token.tag	-0,33	2,62	-1,44
1940_fs_Empty_saddles.txt.token.tag	-1,74	-0,52	1,25
1940_fs_Every_night_about_this_time.txt.token.tag	0,01	-1,30	0,39
1940_fs_Hands_across_the_table.txt.token.tag	1,27	-1,30	0,30
1940_fs_Head_on_my_pillow.txt.token.tag	1,87	-0,59	0,54

quadro 21: Amostra da tabela de resultados de escore padronizado dos componentes fatoriais

<sup>146</sup> Do original: “A factor score is computed by summing, for each text, the number of occurrences of the features having salient loadings on that factor”

<sup>147</sup> Uma tabela com os escores padronizados de todos os componentes fatoriais para os 30 primeiros arquivos do corpus pode ser vista no Anexo 5.

O quadro apresenta as variáveis ‘possessives’, ‘past\_tense’ e ‘present\_tense’ as quais, com escores padronizados, foram renomeadas para ‘Zpossessives’, ‘Zpast\_tense’ e ‘Zpresent\_tense’. O arquivo com todas essas variáveis foi salvo como pat\_mda\_all\_vars\_3.sav.

De posse desses dados, escores padronizados de variáveis por texto, podemos calcular o escore do texto por fator. Para esse cálculo, consideramos as variáveis salientes em cada fator, como apontamos nas seções anteriores, a partir dos valores das matrizes-padrão de cada esfera de análise. O cálculo é feito a partir da somatória dos escores padronizados com polaridade positiva, subtraídos da somatória dos escores padronizados com polaridade negativa (BIBER, 1988, p. 87-91).

O SPSS também realiza esse cálculo, basta selecionar a opção “components” em sua tela principal, seguida por “transform” e “compute variables”. Na janela que surge dessas escolhas, aparece uma coluna com todas as variáveis do arquivo pat\_mda\_all\_vars\_3.sav.

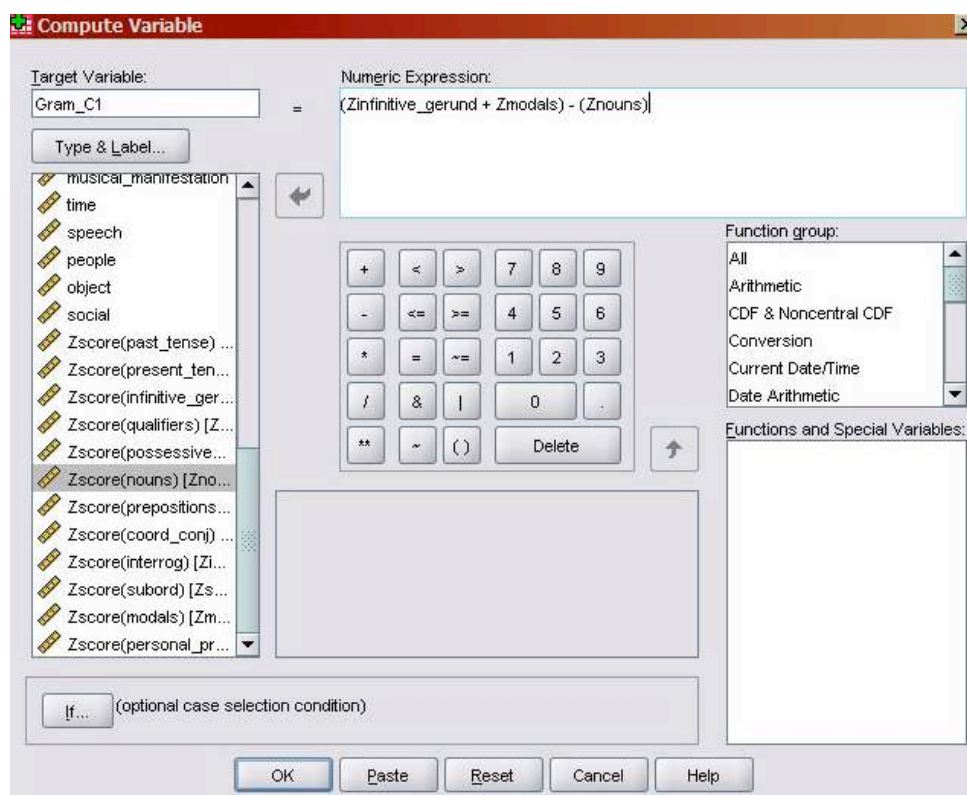


figura 15: Imagem da tela de calculo de variáveis “compute variable” do programa SPSS para as variáveis do fator 1, da instância gramatical

A imagem acima apresenta os procedimentos para o fator 1 da esfera gramatical, denominado Gram\_C1, no campo “*Target Variable*”. No campo “*Numeric Expression*”, colocamos o cálculo que queremos realizar, selecionando as variáveis a partir do campo “*Type & Label*” e as expressões matemáticas presentes na janela. Em seguida, basta clicar em “OK” que o cálculo é realizado para cada linha do arquivo original, ou seja, para cada texto.

O cálculo é feito a partir da matriz fatorial para cada fator, ou seja, são somados os valores das variáveis de mesma polaridade e subtraídos os totais das variáveis de polaridade diferente. No exemplo anterior, do fator gramatical 1, foi consultada a matriz fatorial deste fator (cf. figura 11), que apresentou as variáveis “*infinitive*” (0,636) e “*modals*” com valores positivos (0,615) e a variável “*nouns*” com valor negativo (-0,749)<sup>148</sup>.

Denominamos cada fator com seu próprio número, a partir da esfera de análise de origem. Em outras palavras, para a esfera gramatical, denominamos os fatores Gram\_C1, Gram\_C2 e Gram\_C3; para a esfera semântica denominamos os fatores Sem\_C1, Sem\_C2 e Sem\_C3; e para a esfera léxico-gramatical denominamos os fatores LexGram\_C1, LexGram\_C2 e LexGram\_C3.

Apresentamos nesta seção os passos para a extração de fatores em nosso corpus de estudo, com algumas ilustrações do programa de análise fatorial SPSS. Com isso, encerramos este capítulo de procedimentos metodológicos. A seguir, apresentaremos os resultados da extração de fatores e a análise que nos levou às dimensões linguísticas de letras de música anglo-americanas.

---

<sup>148</sup> As outras variáveis presentes no fator tiveram seus valores ou abaixo do ponto de corte de 0,35 ou valores maiores nos outros componentes e, por isso, não foram consideradas para o cálculo.



## 5. Resultados e Discussão

Neste capítulo apresentaremos os fatores obtidos a partir de nossa análise estatística fatorial de características linguísticas em letras de música anglo-americana. Os fatores serão interpretados quanto aos seus componentes nas três esferas apresentadas no capítulo anterior: **gramatical, semântica e léxico-gramatical**. A interpretação dos fatores deu vazão às dimensões de análise textual de letras de música, que serão apresentadas em seguida. Finalizamos o capítulo apresentando uma discussão dos resultados.

Os resultados obtidos a partir da interpretação das dimensões visam a responder às perguntas de pesquisa apresentadas no capítulo introdutório, as quais serão retomadas ao longo da discussão aqui apresentada. Ainda assim, elas estão reproduzidas a seguir:

1. Quais os principais fatores<sup>149</sup> responsáveis pela variação linguística presente nas letras de música popular anglo-americana?
2. Quais dimensões (funções comunicativas) expressam os fatores obtidos?
3. Como essas dimensões são representadas em relação aos diferentes gêneros musicais presentes em nosso corpus de estudo?
4. Como essas dimensões são representadas em relação aos diferentes artistas selecionados para o estudo?
5. Como essas dimensões são representadas ao longo do tempo por nós estudado (1940-2009)?

Já foi dito que o nosso estudo está baseado em um corpus composto por 6.291 letras de música escritas e gravadas por 32 artistas diferentes, falantes nativos de língua inglesa (americanos, ingleses e canadenses). A fim de que nossos dados fossem quantificáveis para serem analisados na forma de fatores estatísticos, o corpus foi etiquetado por suas características linguísticas de cunho gramatical (morfofossintáticas) e

---

<sup>149</sup> Agrupamentos de variáveis (características linguísticas) coocorrentes, como dissemos anteriormente

semântico. Também foram consideradas para a análise os 200 trigramas mais frequentes em nosso corpus de estudo e a frequência dos mesmos trigramas no inglês geral online, a partir do corpus Google N-Grams.

Os procedimentos preliminares de análise fatorial, conforme expostos no capítulo anterior, apontaram para uma análise em três esferas diferentes, de acordo com a tipologia de suas características. A extração fatorial apontou três componentes fatoriais para cada uma das esferas, como apresentamos pelos gráficos *scree* e pelas tabelas de matrizes fatoriais no capítulo anterior. A seguir, apresentamos cada um desses fatores e sua interpretação.

### 5.1. Interpretação dos fatores

Como vimos no final do capítulo metodológico, a última fase para a extração de fatores envolve um cálculo dos escores padronizados para cada componente. Em nossa análise, foram desenvolvidos nove cálculos diferentes, pois cada esfera de análise (gramatical, semântica e léxico-gramatical) apresentou três fatores. Os cálculos são feitos primeiramente a partir da observação da matriz fatorial de cada esfera. Nela são observados os componentes com valores mais altos (superiores a 0,35)<sup>150</sup> e as diferentes polaridades. Dessa forma, são somados os valores de mesma polaridade e, em seguida subtraídos dos de polaridade diversa.

Cada fator foi intitulado de acordo com a esfera de análise, ou seja, a análise dos componentes gramaticais resultou nos fatores **Gram**, a análise dos componentes semânticos resultou nos fatores **Sem** e a análise dos componentes léxico-gramaticais resultou nos fatores **LexGram**.

A matriz padrão, para as variáveis **gramaticais**<sup>151</sup>, retornou 3 componentes cujos valores acima do ponto de corte estão dispostos a seguir:

---

<sup>150</sup> Conforme explicamos no capítulo metodológico, estipulamos o valor absoluto de 0,35 como mínimo para a análise, conforme recomendado por Biber (1988, p.93).

<sup>151</sup> Como nossa análise é feita em três esferas, optamos por destacar o início da explicação para cada uma delas em negrito, a fim de evitarmos uma nova subdivisão no texto.

<b>variables</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>
nouns	- 0,749		
infinitive_gerund	0,636		
modals	0,615		
personal_pronouns	- 0,401	0,775	
qualifiers		-0,729	
possessives	0,364	0,719	
past_tense			0,819
present_tense			- 0,747

Quadro 22: reprodução da matriz fatorial para as características gramaticais consideradas

Para cada variável do quadro anterior foram computados os valores salientes. Os pronomes pessoais e os possessivos no componente 1 (C1) não são computados, porque apresentam valores maiores no componente 2 (C2), independente de polaridade. Assim, o cálculo para os componentes gramaticais foi feito da seguinte forma:

$$\text{Gram\_C1} = (\text{Zinfinitive\_gerund} + \text{Zmodals}) - (\text{Znouns})$$

$$\text{Gram\_C2} = (\text{Zpersonal\_pron} + \text{Zpossessives}) - (\text{Zqualifiers})$$

$$\text{Gram\_C3} = (\text{Zpast\_tense}) - (\text{Zpresent\_tense})$$

A matriz padrão, para as variáveis **semânticas**, retornou também 3 componentes cujos valores acima do ponto de corte estão dispostos a seguir:

<b>variables</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>
movement	0,632		
emotion		0,759	
musical_manifest.			0,865
time	0,454		-0,366
speech	0,532		
people	0,534		
object	0,494		
social		0,769	

Quadro 23: reprodução da matriz fatorial para as características semânticas consideradas

Assim como para as características gramaticais, as características semânticas também tiveram seu escore calculado considerando os valores mais salientes em cada componente. Dessa forma, a característica “time” (tempo) não foi considerada para o cálculo do componente 3 (C3), mas sim para o componente 1 (C1). Como não há variáveis com polaridade negativa, o cálculo para os componentes semânticos foi feito apenas pela soma das variáveis no polo positivo, não apresentando oposição de

polaridade para esses fatores. O cálculo pode ser melhor visualizado a seguir:

$$\text{Sem\_C1} = Z_{\text{movement}} + Z_{\text{time}} + Z_{\text{speech}} + Z_{\text{people}} + Z_{\text{object}}$$

$$\text{Sem\_C2} = Z_{\text{emotion}} + Z_{\text{social}}$$

$$\text{Sem\_C3} = Z_{\text{musical\_manifestation}}$$

A matriz padrão, para as variáveis **léxico-gramaticais**, retornou também 3 componentes cujos valores acima do ponto de corte estão dispostos a seguir:

<b>variables</b>	<b>C1</b>	<b>C2</b>	<b>C3</b>
nouns	- 0,792		
infinitive_gerund	0,464		
modals	0,442		
personal_pronouns	- 0,373	0,813	
qualifiers		- 0,685	
possessives	0,460	0,649	
past_tense			0,816
present_tense			- 0,706
song_bundles	0,482		
google counts	0,722		

Quadro 24: reprodução da matriz fatorial para as características léxico-gramaticais consideradas

Da mesma forma como visto anteriormente, apenas foram consideradas, para a extração de escores, as variáveis cujos valores foram superiores ao ponto de corte de 0,35<sup>152</sup> e os maiores por variável em relação aos três componentes, sendo desprezados, portanto, para o componente 1 (C1) as variáveis “*personal\_pronouns*” e “*possessives*”. O cálculo pode ser melhor observado pelas expressões a seguir:

$$\text{LexGram\_C1} = (Z_{\text{infinitive\_gerund}} + Z_{\text{modals}} + Z_{\text{songbundles}} + Z_{\text{google\_count}}) - (Z_{\text{noun}})$$

$$\text{LexGram\_C2} = (Z_{\text{personal\_pron}} + Z_{\text{possessives}}) - (Z_{\text{qualifiers}})$$

$$\text{LexGram\_C3} = (Z_{\text{past\_tense}}) - (Z_{\text{present\_tense}})$$

Para cada um desses fatores o programa SPSS criou uma nova coluna na planilha original, contendo o escore para cada texto. A título de ilustração, reproduzimos uma imagem com as dez primeiras linhas dessa planilha a seguir:

<sup>152</sup> Conforme recomendado por Biber (1988, p. 93).

1	file	Gram_C_1	Gram_C_2	Gram_C_3	Gram_C_4	Sem_C_1	Sem_C_2	Sem_C_3	LexGram_1	LexGram_2	LexGram_3
2	1940_fs_April_played_a_fiddle.txt.token	- 0,85	0,82	7,32	- 0,85	- 1,65	- 1,99	2,62	- 0,93	0,82	4,06
3	1940_fs_Empty_saddles.txt.token.tag	- 0,42	- 2,99	- 2,84	- 0,42	0,81	- 1,96	- 0,37	- 2,87	- 2,99	- 1,77
4	1940_fs_Every_night_about_this_time.l	- 1,59	- 0,69	- 2,10	- 1,59	0,31	3,99	2,37	- 1,66	- 0,69	- 1,69
5	1940_fs_Hands_across_the_table.txt.tc	0,04	3,42	- 2,05	0,04	0,59	- 1,07	- 0,37	2,73	3,42	- 1,60
6	1940_fs_Head_on_my_pillow.txt.token.	- 1,26	7,33	- 1,47	- 1,26	- 0,63	- 0,17	- 0,37	- 1,79	7,33	- 1,13
7	1940_fs_ILL_Take_taluhlah.txt.token.ta	2,97	- 0,36	- 1,87	2,97	0,20	- 1,13	- 0,02	1,80	- 0,36	- 0,73
8	1940_fs_ILL_never_let_a_day_pass_by	2,08	- 0,62	2,41	2,08	- 0,27	1,66	0,11	1,39	- 0,62	1,41
9	1940_fs_ILL_never_smile_again.txt.tok	4,18	2,65	- 3,58	4,18	- 1,64	3,05	- 0,37	4,89	2,65	- 2,08
10	1940_fs_ILL_string_along_with_you.txt	1,16	- 1,76	- 0,44	1,16	- 0,43	0,59	0,07	1,47	- 1,76	- 0,77

figura 16: Imagem da tabela do Excel com o valor de cada fator calculado para cada um dos textos do corpus

A imagem anterior mostra que cada texto, ou letra de música, de nosso corpus, recebeu um valor que representa as variáveis mais salientes em cada fator, para cada uma delas. Dessa forma, é possível observar quais textos (ou letras de música, no nosso caso) são mais representativas ou menos representativas de cada fator. Ou seja, quanto maior o valor positivo, mais representativo dessa polaridade o texto (ou letra de música, no nosso caso) deve ser para aquele fator. Quanto maior o valor negativo, mais representativo o texto será da polaridade negativa para aquele fator. Para observar essa diferença, basta ordenar a tabela por ordem crescente para cada coluna que representa o fator que se quer observar.

A interpretação dos fatores apontou, portanto, as características linguísticas predominantes em cada fator e as funções comunicativas por elas desempenhadas, as quais devem estar demonstradas pelas letras de música mais salientes em cada polaridade. Passaremos, portanto, a discutir as características de cada um dos fatores extraídos por nossa análise estatística.

### 5.1.1. Interpretando o Fator Gramatical Gram\_C1

Baseando-se nos valores das variáveis salientes, podemos dizer que o Fator Gram\_C1 é representado pelo contraste entre substantivos e infinitivos/gerúndios/modais. Ou seja, textos que apresentem a presença marcante de substantivos devem apresentar baixa ocorrência de infinitivos, gerúndios e modais, e vice-versa. Segundo Kauffmann (2005, p. 97).

Uma das maneiras de se “adquirir conhecimentos sobre os fatores é avaliar os textos que estão nos extremos dos escores dos fatores” e os que apresentam escores com

valores próximos intermediários (KAUFFMANN, 2005, p. 97). Os textos em posição extrema devem indicar a diferença de polaridade entre as características, enquanto um texto com escore médio deve marcar “a posição ‘neutra’ da manifestação dos fatores” (id.).

A letra de música mais representativa deste fator, ou seja, com valor fatorial mais alto no polo positivo em nosso corpus foi “*Should I*”, gravada por Frank Sinatra em 1950, com **escore 13,966**, reproduzida a seguir:

<p>Should I reveal exactly how I feel?  Should I confess I love you?  Should I recite beneath the pale moonlight?  And swear by the stars above you?  Could I repeat the sweetest story told?  Could I entreat, would it be too bold?  Should I reveal exactly how I feel?  Should I confess I love you?</p> <p>Could I repeat that sweetest story told?  Could I entreat, would it be too bold?  Should I reveal exactly how I feel?  Should I confess I love you?  And swear by stars above you?  Should I confess I love you?</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 1: Letra da música *Should I* – gravação de Frank Sinatra, 1950

A letra dessa música, de fato, apresenta grande concentração de modais (*should*, *could*) e dos verbos no infinitivo que os acompanham (*reveal*, *confess*, *recite*, *swear*, *repeat*, *entreat*, *be*), 13 ocorrências **somadas**. Em contrapartida, apresenta baixo índice de substantivos, apenas 5 ocorrências (das formas *moonlight*, *stars*, *story*), para um texto com 95 palavras ao todo.

A letra de música com maior escore no polo negativo para esse fator é uma gravação de 1998, da cantora Madonna, “*Candy Perfume Girl*”, com **escore -8,024**. Essa letra deve apresentar comportamento oposto à apresentada anteriormente. Vejamos:

<p>Young velvet porcelain boy  Devour me when you're with me  Blue wish window seas  Speak delicious fires</p> <p>I'm your candy perfume girl  Your candy perfume girl</p> <p>Moist warm desire  Fly to me</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

I'm your candy perfume girl  
 Your candy perfume girl  
 I'm your candy perfume girl  
 Candy, candy  
  
 Rush me ghost you see  
 Every center my home  
 Fever steam girl  
 Throb the oceans  
  
 Your candy perfume girl  
 Your candy perfume girl  
 Your candy perfume girl  
 Candy perfume girl  
  
 Did I lie to you?  
 Candy perfume girl  
 Did I lie to you?  
 Magic poison  
  
 You're a candy perfume boy  
 A candy perfume boy  
 You're a candy perfume boy  
 You're candy  
  
 Candy perfume girl  
  
 The sacred nerve is magic poison  
 It's candy, it's candy  
  
 I'm your candy perfume girl  
 I'm your candy perfume girl  
 Boy, girl, boy, girl, boy  
 Candy

Texto 2: Letra da música *Candy perfume girl* – gravação de Madonna, 1998

A letra da música apresenta apenas 5 verbos no imperativo e nenhum modalizador, em oposição a 43 substantivos, em 138 palavras.

Como vemos, a presença marcante das variáveis positivas ou negativas não implica ausência das de polo oposto, mas sua baixa frequência. Assim sendo, uma música que recebe um escore intermediário deve apresentar um equilíbrio de ocorrência entre as variáveis. Esse é o caso da música “*Yours, Sincerely*”, gravada por Ray Charles em 2003, cujo **escore** foi **1,997**. Vejamos:

Dearest one  
 I write what I'm afraid to speak,  
 I'm weak when I'm with you,  
 Tears of love are causing  
 All the ink to blot  
 So what am I to do?  
 Hoping to find the phrases,  
 Groping to find each word,  
 How they all burn like blazes,  
 Now they all seem absurd.  
 Though I don't know where to end  
 And to begin  
 You must give in because

I'm yours sincerely,  
 The one who loves you dearly,  
 To think about it nearly takes  
 My breath away.  
 Very truly  
 My passion is unruly,  
 A dream of you is newly born  
 Each night and day.  
 Oh but my thoughts are fervent,  
 How can I make them plain?  
 Ever your humble servant  
 Faithfully I remain.  
 I'm intending  
 To find a happy ending  
 Because I love you dearly  
 I'm sincerely yours.

Texto 3: Letra da música *Yours, sincerely* – gravação de Ray Charles, 2003

Essa letra contém 136 palavras e apresenta 15 ocorrências de substantivos, entre os quais “*love, ink, phrases, word, blazes, breath, ending*”; em oposição a 18 ocorrências de verbos no imperativo, infinitivo, gerúndio e modal, tais como “*speak, causing, find, end, can, must*”<sup>153</sup>.

As três letras de música representam a variação no fator um entre o escore mais positivo, o médio e o mais negativo, no que diz respeito à frequência concomitante de substantivos, infinitivos, gerúndios e modais. Em outras palavras, o fator Gram\_C1 é marcado pela oposição entre o polo verbal de cunho integrador – marcado pelos verbos no infinitivo, gerúndios e modais – e o polo nominal (marcado pelos substantivos). Verbos no infinitivo geralmente são utilizados para integrar o texto, no caso do texto 1, eles acompanham os modais “*should*” e “*could*” unificando o texto. Já os substantivos indicam informações e são indicativos de grande densidade de informação, ou seja, são

<sup>153</sup> No processo de etiquetagem automática, palavras como *can* e *must* foram categorizadas como modais.



normalmente utilizados em situações que requerem um foco informacional e considerada escolha lexical, como no caso do texto 2, onde a autora representa de forma concreta (por meio dos substantivos) o status da relação amorosa entre os participantes.

Podemos inferir, portanto, que este fator é caracterizado por um discurso do tipo persuasivo, seja do tipo argumentativo, pelo uso de modais, infinitivos e gerúndio, ou por um discurso informativo, pelo uso de substantivos.

### 5.1.2. Interpretando o Fator Gramatical Gram\_C2

O Fator Gram\_C2 é representado pelo contraste entre pronomes pessoais + possessivos com qualificadores. Uma letra de música representativa deste fator, no polo positivo, deve apresentar forte ocorrência desses tipos de pronomes e baixa ocorrência de qualificadores. Vejamos a mais representativa do polo positivo, com **escore 18,311**, “*Forever*”, gravada pelo conjunto Beach Boys, em 1970.

```

If every word I said
Could make you laugh
I'd talk forever (together my love)
I ask the sky just what we had
Mmm It shone forever (together my love my my my my my my my my my my my my
my my my
my my)
If the song I sing to you
Could fill your heart with joy
I'd sing forever (together my love my my my my)

```

Texto 4: Letra da música *Forever* – gravação dos Beach Boys, 1970

A letra da música, que tem 66 palavras ao todo, aparece recheada de exemplos pronominais (26 ocorrências), especialmente pela constante repetição do possessivo “*my*”. Em oposição a apenas 6 ocorrências de qualificadores, que se expressam na forma dos determinantes “*the*” e “*every*”, e dos advérbios “*just*” e “*forever*”. Nitidamente, não se trata de um discurso elaborado, com expansão de referências, atributos típicos de qualificadores. Por outro lado, a repetição de palavras parece apontar para essa característica típica de letras de música. Ou seja, as letras de música, de maneira geral, se caracterizam por eventos repetitivos, pois esses contribuem para a musicalidade sonora das palavras e também em virtude do refrão, como vimos no capítulo 1.

É importante ressaltar que em nosso corpus, procuramos manter as repetições presentes nas letras de música, considerando ser a forma como a letra é escutada e reproduzida pelos falantes, naturalmente, e, sobretudo, como ela é representada graficamente por seus autores.

O polo oposto pode ser exemplificado pela canção “*Straight Cold Player*”, de 1998, gravada pelo artista Lenny Kravitz, que obteve o maior **escore** negativo **-9,727**. A letra da música é uma repetição constante de seu título, possuindo apenas um pronome “*I*”, que ocorre 5 vezes. A combinação adjetiva “*straight cold*”<sup>154</sup>, mais o determinante “*a*” contribuem para a presença marcante de qualificadores em 26 vezes, numa canção com 77 palavras. Melhor visto quando observada a letra:

```
I'm a straight cold player
Just a straight cold player
Straight cold prayer
Ouah !

Just a straight cold player
Just a straight cold player
Straight cold player
Ouah!

I'm a straight cold player
I'm just a straight cold player
Straight cold player
I'm just a straight cold player, player
Straight cold player
Just a straight cold player
Straight cold player
I'm just a straight cold player
Straight cold player
Just a straight cold player
Straight cold player
```

Texto 5: Letra da música *Straight cold player* – gravação de Lenny Kravitz, 1998

Um exemplo de canção com escore intermediário, onde as variáveis aparecem de forma mais equilibrada é “*If I had three wishes*”, gravada por Frank Sinatra em 1955. A letra tem 95 palavras, das quais 22 representam pronomes possessivos e pessoais (*my, I, me, etc.*) e 24 representam qualificadores, como “*the, two, other, dear, etc.*”.

<sup>154</sup> É preciso salientar aqui que, embora expressões como “*straight cold*” possam ser consideradas como um pacote lexical (*bundle*), neste momento da análise, as palavras são observadas individualmente, por sua categorização morfossintática.

Oh, if I had three wishes  
 my first would be for you  
 And if you became my love  
 I'd return the other two

Oh, if I had three wishes  
 I'd ask for you alone  
 All the other things I want  
 will be mine when you're my own

But if you chose me, that  
 I must use the other two,  
 It wouldn't matter, dear,  
 cause here's what I would do

If I had those three wishes  
 my first would be for you  
 And the second and the third would be  
 that my first wish comes true.

Texto 6: Letra da música *If I had three wishes* – gravação de Frank Sinatra, 1955

Assim como para o fator anterior, as três letras de músicas apresentadas aqui representam o arco de variação entre polaridades, neste caso, entre pronomes pessoais e possessivos, num extremo, e qualificadores em outro extremo. O uso de pronomes pessoais marca o envolvimento das pessoas participantes do discurso, seja diretamente (primeiras e segundas pessoas) ou indiretamente (terceiras pessoas, de quem se fala). Os pronomes possessivos caracterizam um discurso que especifica a posse e o egocentrismo, como é o caso do texto 4, que caracteriza a preocupação do autor com o amor dele (“*my my my love*”). Já o uso de qualificadores indica a presença de marcas emocionais e atributivas, ou seja, de um discurso mais descritivo, como podemos perceber no texto 5, que se presta à descrição do tipo de jogador que é o autor (embora também denote um abuso de repetições características de letras de música).

Enfim, podemos inferir que este fator se caracteriza pela oposição entre um discurso interativo, com alta frequência de pronomes pessoais e possessivos, e um discurso descritivo, pela abundância de qualificadores.

### 5.1.3. Interpretando o Fator Gramatical Gram\_C3

A ocorrência de formas verbais no tempo passado em oposição ao tempo presente é

a característica deste fator. A canção “*Whatever Happened to Christmas*”, gravada em 1968 por Frank Sinatra, é a mais representativa do polo positivo. Ou seja, deve conter mais verbos no passado. Vejamos:

Whatever happened to Christmas?  
 It's gone and left no traces,  
 Whatever happened to the faces  
 or the glow,  
 Whatever happened to Christmas,  
 to Christmas way of living?  
 Whatever happened to the giving,  
 the magic in the snow?

Remember the sight and the smell and the sound,  
 And remember hearing the call,  
 Remember how love was all around,  
 whatever happened to it all?

Whatever happened to Christmas,  
 bells in the streets were ringing,  
 Whatever happened to the singing,  
 the songs we used to know.  
 Whatever happened to this Christmas,  
 and when did it disappear from view,

Where was I, and whatever happened to you?  
 Whatever happened to Christmas and you?

Texto 7: Letra da música *Whatever happened to Christmas* – gravação de Frank Sinatra, 1968

A letra contém 109 palavras das quais 18 são verbos no tempo passado e 4 são verbos no presente. O **escore** desta letra de música é **10,767**. Já o polo oposto deve conter a situação inversa, ou seja, mais verbos no presente que no passado, o que pode ser observado na música “*Because*”, gravada pelos Beatles em 1969.

Ah, because the world is round  
 it turns me on  
 Because the world is round

Ah, because the wind is high  
 it blows my mind  
 Because the wind is high

Ah, love is old, love is new  
 Love is all, love is you

Because the sky is blue  
 it makes me cry  
 Because the sky is blue

Ah, ah, ah, ah.

Texto 8: Letra da música *Because* – gravação dos Beatles, 1969

“*Because*” não contém nenhum verbo no passado e seu **escore** neste fator é o maior de polo negativo **-6,515**. A letra da música é descritiva, caracterizada pelo uso do verbo ser (“*is*”), ou seja, o autor descreve fenômenos da natureza e suas consequências para ele.

A canção no quadro abaixo apresenta um **escore** médio de **1,445**, devendo, portanto, apresentar equilíbrio entre passado e presente. Vejamos:

<p>Heavy hung the canopy of blue          Shade my eyes and I can see you          White is the light that shines through the dress that you wore</p> <p>She lay in the shadow of the wave          Hazy were the visions of her playing          Sunlight on her eyes but moonshine made her cry ev'ry time</p> <p>Green is the colour of her kind          Quickness of the eye deceives the mind          Envy is the bond between the hopeful and the damned</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 9: Letra da música *Green is the color* – gravação de Pink Floyd, 1969

A canção foi gravada pelo conjunto britânico Pink Floyd em 1969. Intitulada “*Green is the Color*”, possui 76 palavras, 5 verbos no passado e 7 no presente. Nessa letra de música, o autor descreve as consequências para o presente de “visões” e “eventos” do passado.

Como na interpretação dos fatores anteriores, as três letras de música ilustram a variação entre os polos opostos que caracterizam os fatores, neste caso, entre passado e presente. Verbos no presente são usados para destacar referenciais imediatos, normalmente associados a um discurso não planejado. O uso de verbos no passado caracteriza uma produção narrativa ou descritiva com foco temporal no passado. Dessa forma, podemos inferir que este fator é caracterizado pela oposição de um discurso narrativo e um discurso com referencial imediato.

Apresentada a interpretação das características dos fatores gramaticais, passamos aos semânticos.

#### 5.1.4. Interpretando o Fator Semântico Sem\_C1

A análise fatorial de cunho semântico levou em consideração grupos de palavras que se associam sob um mesmo tema. Na esfera semântica não houve contraste de polaridade para nenhum dos fatores. Isso significa que a presença das variáveis mais salientes de cada fator não implica na ausência ou menor presença de outras variáveis específicas. Na realidade, os textos (letras de música) deverão ser caracterizados pela abundância do conjunto de variáveis de cada fator.

O fator Sem\_C1 apresenta a maior combinação de variáveis semânticas, as quais devem indicar movimento, tempo, fala, pessoas e objetos.

Vejamos como se apresenta a letra de música com o maior escore neste fator:

```
Birds fly up
And down spinnin' round
Flyin' all around
From my window
A little brown sparrow came
Flutterin' down

Le moineau est venu
Se poser ma fenetre

He came to my window

And as I watched him
I found myself reelin' and turnin' around
And the faster I ran
I ran out of breath and I fell to the ground
(Poof)

Ahhhhhhhhhhhh

He came to my window
He came to my window
He came to my window

From their eyes
The people must look like miniature toys

Do do do do do
Do do do do do do
Fly away
Fly away
```

Texto 10: Letra da música *At my window* – gravação dos Beach Boys, 1970

A canção “*At my window*”, gravada pelo conjunto americano Beach Boys apresenta um **escore de 42,475** para este fator. Ela o representa especialmente pela presença de

palavras como “*fly*”, “*spinning*”, “*came*”, “*ran*”, “*fell*” e “*look*” que indicam movimento e aparecem concomitantemente com palavras que indicam objeto como “*window*”, “*eyes*”, “*toys*” e “*ground*”, e com palavras que indicam pessoas como “*I*” e “*people*”.

Outro exemplo dessa categoria é a canção “*Mountain Greenery*”, gravada por Frank Sinatra em 1963. Essa canção recebeu o **escore** de **34,749**. Vejamos:

<p>In a mountain greenery, where God paints the scenery,          Just two crazy people together,          While you love your lover, let blue skies be your coverlet          When it rains we'll laugh at the weather.          And if you're good, I'll search for wood,          So you can cook, while I stand looking.          Beans could get no keener reception in a beanery,          Bless our mountain greenery home!</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 11: Letra da música *Mountain Greenery* – gravação de Frank Sinatra, 1963

Podemos também observar na letra desta canção a forte presença de indicadores de objeto como “*mountain*”, “*scenery*”, “*skies*”, “*wood*”, “*beans*”, etc; e indicadores de pessoas como “*people*”; a presença de ações que indicam fala, tais como “*let*” e “*bless*”; a presença de indicadores de movimento como “*paints*”, “*rains*”, “*cook*” e “*get*”.

A música “*Neiani*”, gravada por Frank Sinatra em 1940, foi a música que recebeu menor **escore** neste fator, **-3,436**. Isso não significa que a letra não contenha nenhuma das características deste fator, mas que a ocorrência simultânea das características aconteça em menor frequência que nas anteriores. Vejamos:

<p>I found her in Hawaii, my Neiani, so lovely          I loved her in Hawaii, my Neiani, my own.          I remember soft nights, love songs, moonlight,          Then I sailed away and left my sweet Neiani in Hawaii of old.</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 12: Letra da música *Neiani* – gravação de Frank Sinatra, 1940

A letra da música é bem pequena, apenas 38 palavras. Encontramos demonstrativos verbais que indicam movimento como “*found*” e “*sailed*” que têm suas frequências equilibradas por presenças verbais que indicam ações mentais, como “*loved*” e “*remembered*”. Não há menção de tempo, nem de fala. Palavras como “*Neiani*” e “*Hawaii*” fazem parte do fator, mas a música apresenta alta frequência de qualificadores (“*lovely*”, “*own*”, “*soft*”) o que ofusca a concomitância das características deste fator.

As letras das duas músicas apresentadas aqui representam a presença conjunta das características semânticas para o fator semântico C1, ou seja, alta frequência concomitante de palavras que indicam movimento, tempo, fala, pessoas e objetos. Isso nos parece um indicativo de letras de música que tratam de ações pessoais relacionadas ao momento em que ocorreram e aos objetos e pessoas envolvidos na ação.

### 5.1.5. Interpretando o Fator Semântico Sem\_C2

Este fator também não apresenta diferença de polaridade, mas é marcado pela presença simultânea de palavras que indicam emoção e manifestação/problemas de ordem social. Melhor compreendido quando observamos a letra de uma música.

```

You never give me your money
Your only give me your funny paper
And in the middle of negotiations you break down

I never give you my number
I only give you my situation
And in the middle of investigation I break down.

Out of college money spent
See no future pay no rent.
All the money's gone, nowhere to go.

Any Jobber got the sack,
Monday morning turning back.
Yellow lorry slow, nowhere to go.
But oh - that magic feeling nowhere to go.
One sweet dream

Pick up the bags and get in the limousine.
Soon we'll be away from here.
Step on the gas and wipe that tear away,

One sweet dream came true today, came true today.
One, two, three, four, five, six, seven,
All good children go to heaven.

```

Texto 13: Letra da música *You never give me your money* – gravação dos Beatles, 1969

A música “*You never give me your money*”, gravada em 1969 pelos Beatles, tem **escore 10,948**. Entre suas 133 palavras estão muitas que representam manifestações de cunho social, ou seja, palavras relacionadas com a experiência social das pessoas, tais como “*money*”, “*paper*”, “*negociations*”, “*situation*”, “*investigation*” e “*college*”, entre outras. Além de outras palavras que podem ser associadas a emoções, tais como



“future”, “feeling”, “dream”, “magic” e “tear”.

A música exposta a seguir, “*Burning man*”, da banda Queensryche, gravada em 1999, é a que possui o maior **escore** negativo, **-2,563**. Vejamos por que:

<p>Miles of desert fly, the quiet of dawn. I ride the winds of morning to see the man that's burning.</p> <p>Burning Man</p> <p>Miles of desert fly, the quiet of dawn. I ride the winds of morning to see the man that's burning.</p> <p>In my head I hear the song a million voices strong. I feel the heart of conscious returning, that's why the man is burning.</p> <p>Burning Man</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 14: Letra da música *Burning man* – gravação de Queensryche, 1999

Não podemos encontrar palavras que estejam relacionadas a questões sociais nesta letra de música. As palavras “feel”, “heart” e “conscious” representam emoções. Todavia, o fator é caracterizado pela presença concomitante das duas variáveis, o que não ocorre aqui, daí seu escore negativo.

Podemos observar, portanto, que as duas variáveis semânticas, emoção e manifestação/problemas de ordem social, podem tipicamente aparecer em conjunto, representando o segundo fator semântico, o que indica uma tendência temática a letras de música que falam sobre emoção relacionada à sociedade.

### 5.1.6. Interpretando o Fator Semântico Sem\_C3

O terceiro fator semântico é representado pela presença de palavras associadas à manifestação musical, o que deve envolver música, dança, canto e som, inclusive os participantes desses processos.

Para ilustrar este fator, selecionamos a canção “*Rock’n’roll music*”, gravada pelos Beach Boys em 1976, cujo **escore** positivo foi de **5,608**. Cabe aqui esclarecer que a música foi composta originalmente por Chuck Berry e gravada pelo compositor em 1957. O artista não integrou nosso corpus de estudo, mas sua canção apareceu mais de

uma vez na seleção preliminar do corpus, pois foi gravada também pelos Beatles, em 1963. Como explicamos no início do capítulo, quando o corpus apresentava a letra da mesma canção mais de uma vez por ter sido gravada por artistas diferentes, mantivemos apenas uma versão. No caso desta, ficou arquivada a gravação dos americanos Beach Boys.

just let me hear some of that rock and roll music  
 Any old way you choose it  
 It's got a backbeat you can't lose it  
 Any old time you use it  
 It's gotta be rock and roll music  
 If you wanna dance with me  
 If you wanna dance with me

I have no kicks against modern jazz  
 Unless they try to play it too darned fast  
 And change the beauty of the melody  
 Until it sound just like a symphony

That's why I go for that rock and roll music  
 Any old way you choose it  
 It's got a backbeat you can't lose it  
 Any old time you use it  
 It's gotta be rock and roll music  
 If you wanna dance with me  
 If you wanna dance with me

I took my loved one over 'cross the tracks  
 So she could hear my man a-wailin' sax  
 I must admit they have a rockin' band  
 Man they was blowing like a hurricane

That's why I go for that rock and roll music  
 Any old way you choose it  
 It's got a backbeat you can't lose it  
 Any old time you use it  
 It's gotta be rock and roll music  
 If you wanna dance with me  
 If you wanna dance with me

Way down south they gave a jubilee  
 I tell you folks they had a jamboree  
 And drinking beer from a wooden cup  
 The folks a-dancin' got all shook up

They started playing that rock and roll music  
 Any old way you choose it  
 It's got a backbeat you can't lose it  
 Any old way you use it  
 It's gotta be rock and roll music  
 If you wanna dance with me  
 If you wanna dance with me

Say if you wanna dance with me

<p>If you wanna dance with me          If you wanna dance with me          It's gotta be some of that          If you wanna dance with me</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 15: Letra da música *Rock'n'Roll music* – gravação de Beach Boys, 1976

A letra da música é uma representação típica deste fator e fala sobre música e dança o tempo todo: “*rock*”, “*music*”, “*dance*”, “*track*”, “*sax*”, “*play*”, “*jazz*”, “*melody*”, “*symphony*”, “*backbeat*”, etc.

Uma canção com alto valor na polaridade negativa deve apresentar pouca representatividade de palavras que sugiram música, dança, som e ritmo. Vejamos a letra de “*My Bonnie*”, conforme foi gravada pelos Beatles, que recebeu o mais alto **escore** negativo **-0,365**.

<p>My Bonnie lies over the ocean,          my Bonnie lies over the sea.          My Bonnie lies over the ocean.          Oh bring back my Bonnie to me.</p> <p>My Bonnie lies over the ocean,          my Bonnie lies over the sea.          Well my Bonnie lies over the ocean.          Yeah bring back my Bonnie to me.</p> <p>Yeah bring back, ah bring back,          oh bring back my Bonnie to me to me.          Oh bring back, oh bring back,          oh bring back my Bonnie to me.</p> <p>Well my Bonnie lies over the ocean,          my Bonnie lies over the sea.          Yeah my Bonnie lies over the ocean.          Oh I said bring back my Bonnie to me.</p> <p>Yeah bring back, ah bring back,          oh bring back my Bonnie to me to me.          Oh bring back, ah bring back,          oh bring back my Bonnie to me.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 16: Letra da música *My Bonnie* – gravação de Beatles, 1976

Como podemos observar, não há nenhuma palavra que possa ser associada diretamente à música, ritmo, som ou dança, o que justifica sua posição extrema oposta à outra do fator (*Rock and Roll music*). Esse fator sugere, portanto, a existência de letras de música que abordem a temática de manifestações associadas à música.

Explicados os fatores semânticos, passaremos à descrição dos fatores léxico-gramaticais.

### 5.1.7. Interpretando os Fatores Léxico-gramaticais

Diferentemente das interpretações feitas até o momento, para as outras esferas de nossa análise, a esfera léxico-gramatical será interpretada em um único item neste capítulo. Isto ocorre devido à semelhança de seus resultados quando contrastados com os resultados obtidos nos fatores de cunho exclusivamente gramatical, conforme apresentamos no capítulo anterior. Na realidade, embora apresentem leve alteração numérica (ver quadro 20), os fatores LexGram\_C2 e Gram\_C2 apresentam-se ambos caracterizados pela oposição entre pronomes pessoas e possessivos X qualificadores, enquanto os fatores LexGram\_C3 e Gram\_C3 são marcados pela presença verbal opondo passado X presente.

A extração dos escores das letras de música mais ou menos salientes, para as instâncias gramaticais e léxico-gramaticais, retornou os mesmos arquivos, praticamente na mesma ordem para as duas extrações. Por exemplo, a letra da música “*Forever*” usada para ilustrar o fator gramatical Gram\_C2, por ter o maior índice de polaridade positiva, é também a que apresenta a maior polaridade positiva para o fator LexGram\_C2, inclusive com os mesmos escores de **18,311**. O mesmo acontece para a letra da música “*Straight Cold Player*” que é a de maior valor na polaridade negativa (**-9,727**).

No caso do fator LexGram3, que é representado pela oposição entre passado e presente, obtivemos também resultado semelhante. Isto é, embora o valor dos escores tenha sofrido uma leve alteração, as canções mais tipicamente representativas deste fator são as mesmas. A canção “*Because*”, dos Beatles, que representa o polo negativo do fator Gram\_C3 com **escore -6,515**, apresenta **escore -5,023** neste fator. Já a canção selecionada para representar o polo positivo no fator gramatical “*Whatever Happened to Christmas*”, que trazia o maior índice fatorial (**10,767**), perdeu a primeira posição de representatividade neste fator para a gravação de “*Guilty by Association*”, da cantora Madonna. Aqui, a primeira obteve um escore de **4,855**, enquanto a segunda atingiu um escore de **6,659**.

Precisamos ressaltar, contudo, que, apesar da variação de valores, ambas as músicas encontram-se nas posições iniciais, nos dois fatores. Mesmo assim, reproduzimos a seguir a canção mais representativa do fator LexGram\_C2:

```

You've been sanctified
And I've been tried
Guilty by association

You've been canonized
And I've been fried
Guilty by association

All the little loonies
With a salient obsession
Come out from the boonies
With their sharpies and their guns
And we're loaded with questions

Guilty by association
Guilty by association

You've been idolized
And I've been tried
[I've been tried]
Guilty by association

You've been terrorized
And I've been fried
Guilty by association

All the little loonies
With a salient obsession
Come out from the boonies
With their sharpies and their guns
They're all loaded with questions

Loaded with questions
Loaded with questions

```

Texto 17: Letra da música *Guilty by association* – gravação de Madonna, 1994

A canção apresenta forte presença de verbos no passado, como “*been*”, “*sanctified*”, “*tried*”, “*canonized*”, “*fried*” etc., com uma única manifestação de presente, “*come*”. Podemos perceber, portanto, o mesmo tipo de oposição que caracteriza o fator Gram-C3, isto é, a oposição entre um discurso narrativo sequencial com foco no passado e um discurso com referencial imediato.

Assim sendo, nosso foco interpretativo para os componentes léxico-gramaticais foi dado ao componente Lexgram\_C1, pois apresenta certa diferença entre os componentes

apenas gramaticais. Em outras palavras, a nossa análise fatorial de cunho léxico-gramatical incluiu as variáveis gramaticais e as variáveis de contagem de trigramas na música e no corpus do Google N-grams. Estas duas novas variáveis se agruparam no fator 1, que ficou composto pela oposição de polaridade entre as seguintes variáveis:

polaridade positiva	polaridade negativa
infinitivos e gerúndios modais song_bundles google_counts	substantivos

Quadro 25: oposição de polaridade entre os componentes do fator Lexgram\_C1

Quando fizemos a extração para os componentes gramaticais, a oposição de polaridade extraída foi praticamente a mesma, a não ser pela não inclusão dos trigramas (*song bundles e google counts*). Em suma, o fator Lexgram C1 apresenta as mesmas características gramaticais do fator Gram\_C1 acrescidas das contagens de trigramas.

Vale lembrar ao leitor que os trigramas foram inseridos na contagem juntamente com todas as outras variáveis e a extração fatorial os agrupou no componente C1. Isso indica que quando consideramos a sequência de três palavras mais frequentes em nosso corpus de estudo<sup>155</sup>, também aparece em oposição a substantivos individuais, ainda que aqueles representem palavras de conteúdo.

Apesar da inclusão dos trigramas, a diferença entre o fator que ora descrevemos e o fator de cunho exclusivamente gramatical não é tão expressiva assim. A letra da música mais representativa deste fator no polo positivo, com **escore de 14,319** – “*My number one*”, gravada em 2009, pelo grupo americano Paramore –, na esfera gramatical ocupa a terceira posição, apresentando escore de 13,966. Vejamos a letra:

<p>You're my number one          You're my golden star          I look at earth from here          Still you don't seem so far</p> <p>But you will never know (you will never know)          Oh you will never know</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

<sup>155</sup> Foram considerados os 200 trigramas mais frequentes em nosso corpus e a contagem dos mesmos trigramas no inglês geral online, a partir do corpus Google N-grams

Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh

Do you think of me  
 And do you wonder  
 If we could ever be  
 We're dying under

Because we will never know (we will never  
 know)  
 Until I see you show through

Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 I'll

I wish I could tell you how I feel  
 And show you what's inside of me is real  
 I don't know what I'm waiting for  
 Can't explain it anymore  
 Yeah

Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 No I will never admit that I  
 Oh oh oh oh oOoh oh oh  
 Oh I will never admit that I  
 Oh oh oh oh oOoh oh oh

Oh oh oh oh oOoh oh oh  
 And I love  
 Oh oh oh oh oOoh oh oh  
 I love  
 Oh oh oh oh oOoh oh oh  
 I love you

Texto 18: Letra da música *My number one* – gravação de Paramore, 2009

Além da presença marcante de modais e infinitivos, como “*will*”, “*know*” e “*admit*”, a canção apresenta frequência 59 para os song bundles. Os trigramas que aparecem nesta canção que também têm frequência superior a 50.000 no corpus do Google N-Gram são “*admit that I*”, “*and do you*”, “*and I love*”, “*and show you*”; “*because we will*”, “*but you will*”; “*could ever be*”, “*could tell you*”; “*do you think*”,

*“how I feel”, “I could tell”, “I look at”, “I love you”; “I see you”, I will never”, “I wish I”, “if we could”; “inside of me”; “know we will”; “know you will”; “my number one”; “of and me”; “oh and I”, “oh oh oh”; “tell ou how”; “think of me”; “until I see”; “we will never”; “will nver know”; “wish I could”; “wonder if we”; “you know I”; “you think of”; “you will never” e “you wonder if”.*

Observamos também a baixa ocorrência de substantivos (apenas 3: *“one”, “star” e “earth”*) que é a variável com polaridade oposta para este fator. Como podemos ver, este fator indica a tipicidade das canções em relação tanto à linguagem da música quanto a essa linguagem presente no inglês geral, preservando a oposição de discurso do tipo argumentativo a um discurso do tipo informativo.

Julgamos que as amostras usadas para ilustrar a interpretação dos fatores foram suficientes o bastante para não sobrecarregar a leitura deste texto. Ademais, outras canções serão apresentadas posteriormente para ilustrar as dimensões.

Nesta seção apresentamos nossa interpretação dos fatores que emergiram de nossa análise estatística, respondendo nossa primeira questão de pesquisa: Quais os principais fatores responsáveis pela variação linguística presente nas letras de música popular anglo-americana?

A fim de respondermos às nossas outras questões, que envolvem as funções comunicativas representadas por esses fatores, isto é, as dimensões, passaremos, a seguir, à interpretação de cunho funcional, verificando sua relação com os gêneros musicais, entre os diferentes artistas e períodos de tempo presentes em nosso corpus de estudo.

## **5.2. Dimensões de variação na música popular anglo-americana**

Conforme dito anteriormente, a pesquisa descrita aqui partiu do pressuposto de que as letras de música popular quando consideradas por suas características linguísticas devem apresentar variações em dimensões que podem não necessariamente estar em consonância com a classificação a que são submetidas pelas gravadoras e a mídia musical.



Em outras palavras, acreditamos que a variação de estilos musicais, cantores e grupos, entre si e ao longo do tempo, pode ser explicada em parte pela variação linguística das letras, não por meio de uma oposição binária, mas pela existência de multidimensões de variação.

Consideramos também o pressuposto de que um texto não deve ser analisado por apenas uma variável linguística, mas por aquelas que mais, ou menos, o caracterizam. Assim, temos o objetivo de desvendar as dimensões de variação linguística que permitem uma caracterização da música popular anglo-americana a partir da observação de suas letras.

Nosso trabalho esteve amparado teoricamente pela Linguística de Corpus que também embasa a proposta metodológica de análise multidimensional que adotamos; a AMD, de Douglas Biber (1988 et seq.). No capítulo anterior descrevemos os passos da análise estatística fatorial que nos forneceu a possibilidade de analisar a música sob três diferentes esferas linguísticas: gramatical, semântica e léxico-gramatical. Essas três extrações são um aspecto que diferencia nossa pesquisa das demais no âmbito da AMD, pois normalmente é escolhida apenas uma extração, de base gramatical. Em nossa pesquisa, ficou evidente que apenas uma extração não seria suficiente para dar conta da riqueza dos dados, por isso mantivemos as três. Para cada uma dessas esferas, a análise fatorial retornou três fatores, os quais foram interpretados por suas características na seção anterior e serão interpretados por seus aspectos funcionais, ou dimensões de análise, nesta seção.

Letras de músicas representantes de cada dimensão foram interpretadas em dimensões sob três perspectivas: década de gravação, artista e gênero/estilo musical. No capítulo metodológico, ao descrevermos nosso corpus de estudo informamos que os dados foram coletados e armazenados de acordo com a década de gravação de cada música num período que compreende os anos de 1940 a 2009. A partir de tal subdivisão, os dados foram também arquivados de acordo com o artista que gravou cada música, o que pode ser observado no quadro 10, para que, dessa forma, os dados fosse acessados de acordo com seu agrupamento por artistas, década ou gênero musical.

Ao longo do Capítulo 1, apresentamos uma categorização da música popular por gêneros/estilos musicais de acordo com diversas fontes (STARR; WATERMAN, 2007; CHARLTON, 2009; Billboard Charts; Rolling Stone Magazine; allmusic.com e musicimprint.com). Vimos que não existe um consenso absoluto quanto a tal categorização, por isso, decidimos pela classificação mais recorrente para cada banda. A classificação por nós adotada está resumida no quadro a seguir. Um quadro com as características apontadas por fontes diferentes, por banda, em contraste à nossa, pode ser visto no Anexo 6.

<b>ESTILO/GÊNERO</b>	<b>Banda</b>
ALTERNATIVE ROCK	Killers
	Red Hot
COUNTRY	Johnny Cash
COUNTRY SOUL	Ray Charles
GRUNGE	Nirvana
	Pearl Jam
HARD ROCK	Aerosmith
	Bon Jovi
	Def Leppard
HEAVY METAL	Iron Maiden
	Metallica
POP	Elton John
	Madonna
	Michael Jackson
POP (VOCAL)	Frank Sinatra
POP ROCK	Beatles
	U2
POST GRUNGE	Creed
	Foo Fighters
	Nickelback
PROGRESSIVE ROCK	Pink Floyd
	Queensryche
PUNK POP	My Chemical Romance
	Paramore
	Simple Plan
PUNK ROCK	Green Day
ROCK	Led Zeppelin
	Lenny Kravitz
	Queen
	Rolling Stones
ROCK (SURF)	Beach Boys
ROCK & ROLL	Elvis Presley

Quadro 26: resumo da categorização dos estilos musicais por artista, em nosso corpus de estudo

As letras de música foram agrupadas por cada uma das perspectivas apontadas no início da seção (ano, estilo e artista) e dispostas em três novas colunas em nosso arquivo

original e um novo procedimento estatístico foi realizado: a análise descritiva.

A análise descritiva<sup>156</sup> computa os dados de cada um dos componentes de cada fator segundo uma perspectiva de análise previamente selecionada. Em nosso caso, queríamos observá-las de acordo com o estilo (*style*), a década (*year*) ou o artista (*singer\_band*). A título de ilustração, apresentamos a seguir uma amostra da nossa planilha do Excel com as dez primeiras linhas de nosso arquivo, contendo agora as colunas com as classificações por ano, artista e estilo, além dos valores de cada componente de cada fator para cada texto:

1	file	Year	Singer_ba	Style	G_C1	G_C2	G_C3	S_C1	S_C2	S_C3	LG_C1	LG_C2	LG_C3
2	1940_fs_April_played_a_fiddle.txt.tok	4	Frank Si	Pop (voc	-0,85	0,82	7,32	-1,65	-1,99	2,62	-0,93	0,82	4,06
3	1940_fs_Empty_saddles.txt.token.tag	4	Frank Si	Pop (voc	-0,42	2,99	-2,84	0,81	-1,96	-0,37	-2,87	-2,99	-1,77
4	1940_fs_Every_night_about_this_tim	4	Frank Si	Pop (voc	-1,59	-0,69	-2,10	0,31	3,99	2,37	-1,66	-0,69	-1,69
5	1940_fs_Hands_across_the_table.txt	4	Frank Si	Pop (voc	0,04	3,42	-2,05	0,59	-1,07	-0,37	2,73	3,42	-1,60
6	1940_fs_Head_on_my_pillow.txt.toke	4	Frank Si	Pop (voc	-1,26	7,33	-1,47	-0,63	-0,17	-0,37	-1,79	7,33	-1,13
7	1940_fs_ILL_Take_taluhlah.txt.token	4	Frank Si	Pop (voc	2,97	-0,36	-1,87	0,20	-1,13	-0,02	1,80	-0,36	-0,73
8	1940_fs_ILL_never_let_a_day_pass	4	Frank Si	Pop (voc	2,08	-0,62	2,41	-0,27	1,66	0,11	1,39	-0,62	1,41
9	1940_fs_ILL_never_smile_again.txt.tr	4	Frank Si	Pop (voc	4,18	2,65	-3,58	-1,64	3,05	-0,37	4,89	2,65	-2,08
10	1940_fs_ILL_string_along_with_you.t	4	Frank Si	Pop (voc	1,16	-1,76	-0,44	-0,43	0,59	0,07	1,47	-1,76	-0,77

Figura 17: tabela do Excel com as 10 primeiras linhas de nosso arquivo acrescidas das classificações por ano, artista e gênero.

Por meio da observação da planilha do Excel, é possível determinar quais letras de música são mais representativas de cada fator, ao se considerar determinada perspectiva de análise. Em outras palavras, podemos determinar as letras de música mais representativas para cada dimensão (Gram\_C1; Gram\_C2; Gram\_C3; Sem\_C1; Sem\_C2; Sem\_C3; LexGram\_C1; LexGram\_C2 e LexGram\_C3) quando observamos artista, as letras de música mais representativas quando observamos o gênero musical e as letras mais representativas quando consideramos o período de gravação. Dessa forma, podemos caracterizar melhor cada dimensão de análise das letras de música, conforme explicitaremos a seguir.

### 5.2.1. Dimensões Gramaticais

A análise de dimensões gramaticais para letras de músicas anglo-americanas levou em consideração os fatores obtidos a partir de 48 variáveis morfossintáticas, que foram agrupadas em 13 características principais. Conforme descrito na seção anterior, a

<sup>156</sup> Uma imagem da janela de análise descritiva para perspectiva “ano” pode ser observada no Anexo 7

análise fatorial dessas características em nosso corpus de estudo retornou três fatores (Gram\_C1, Gram\_C2 e Gram\_C3) que representam as linhas de análise em forma das dimensões que descrevemos agora.

### 5.2.1.1. Primeira Dimensão Gramatical: Argumentação X Informação

A primeira dimensão gramatical pode ser caracterizada pela oposição entre opinião e referência. Ou seja, as características linguísticas gramaticais observadas pelo fator que originou esta dimensão (infinitivos/gerúndios/imperativos e modais versus substantivos) funcionam juntas para marcar o discurso argumentativo ou descritivo/informativo do enunciador.

Vale ressaltar que os elementos interpretativos desta dimensão são bastante semelhantes aos encontrados por Biber (1988, p. 111, 113, 148-151) para sua quarta dimensão (“*Expressão explícita de persuasão*”), visto que a nossa dimensão é caracterizada pela oposição de substantivos e a de Biber não possui fatores negativos e inclui as subordinativas (elemento mais persuasivo). Segundo o autor, as características linguísticas desta dimensão expressam marcas normalmente encontradas em registros do tipo cartas profissionais e editoriais, em oposição a noticiários (*broadcasts*) e críticas de imprensa. A nossa análise, por sua vez, leva em conta um único gênero textual (letras de música), no que se assemelha à pesquisa de Pacheco de Oliveira (1997), enquanto a análise de Biber, por outro lado, contrastava diversos gêneros dos discursos oral e escrito.

A análise descritiva, realizada em três instâncias, parte de uma tabela gerada pelo programa SPSS, a qual contém os escores médios de cada uma para cada dimensão. Assim, é possível ordená-las de forma que se perceba quais letras de música são mais representativas nesta dimensão.

Os escores extraídos para a análise por **cantor ou banda** para o a dimensão Argumentação X Informação estão reproduzidos no quadro a seguir (quadro 27), que apresenta quatro colunas, a primeira contém o nome do artista; a segunda, “*Mean*”,

apresenta o escore médio de frequência das características por artista (média entre a maior e menor frequência); a terceira apresenta o desvio padrão (*Std. Deviation*), que é uma medida de como a característica se distribui entre os textos; e a última marca o número de textos (letras de música) de cada artista. Os dados foram ordenados pela segunda coluna (escore médio), a fim de facilitar a visualização dos artistas com maior e menor frequência nesta categoria. Vejamos:

**Descriptive Statistics**  
Dependent Variable: Gram\_C\_1

Singer_band	Mean	Std. Deviation	N
Simple P	1,7549	2,32232	49
Paramore	1,5996	1,92685	50
Creed	1,4551	2,05388	60
Nickelba	,9810	2,06059	79
My Chemi	,9287	1,64820	61
Lenny Kr	,8173	2,50687	125
Madonna	,8025	1,95550	167
Beatles	,7267	2,28305	281
Foo Figh	,4917	1,87241	109
Michael	,4783	1,93754	99
Elvis Pr	,3159	2,06701	305
Nirvana	,2633	2,73215	98
Def Lepp	,2081	2,17631	142
Bon Jovi	,1571	1,74444	127
Killers	,1395	1,91188	51
Frank Si	,1363	2,34629	1218
Ray Char	,0916	1,98023	206
Led Zepp	,0539	1,66369	79
Queen	,0346	2,01069	172
Total	,0263	2,14360	6290
Rolling	-,0404	2,11504	430
Queensry	-,1763	1,54018	112
Metallic	-,1955	1,74841	119
Red Hot	-,2192	1,89482	169
Pearl Ja	-,2222	2,20518	85
U2	-,2257	2,17230	168
Greed Da	-,2348	1,95319	49
Beach Bo	-,4902	2,06954	276
Pink Flo	-,5363	2,30069	163
Green Da	-,5419	1,97190	110
Johnny C	-,5477	1,80454	439
Iron Mai	-,5938	2,09109	215
Elton Jo	-,6247	1,74597	325
Aerosmith	-,7745	1,81517	152

Quadro 27: Estatística descritiva para o fator Gram\_C1, em relação a cantor/banda

Ao observarmos esta dimensão por artista, verificamos a maior representatividade

para a banda Simple Plan e a menor representatividade para a banda Aerosmith. Podemos inferir, a partir desses dados, que as letras das músicas da banda Simple Plan são mais ricas em discurso argumentativo que as letras das músicas da banda Aerosmith, que são repletas de referenciais concretos. Um quadro com as músicas mais e menos representativas de cada perspectiva de análise para este fator está exposto no Anexo 8. Vejamos letras de música representativas desta dimensão, começando por “*Grow up*”, gravada pela banda Simple Plan, que é a mais representativa:

This is who I am  
and this is what I like  
GC, Sum and Blink and Mxxp rocking my room  
if your looking for me  
I'll be at the show  
I could never find a better place to go

Until the day I die I promise I won't change  
so you better give up  
I don't wanna be told to grow up  
and I don't wanna change  
I just wanna have fun  
I don't wanna be told to grow up  
and I don't wanna change  
so you better give up  
cause I'm not gonna change  
I don't wanna grow up

I like to stay up late  
spend hours on the phone  
hanging out with all my friends  
and never being at home  
I'm impolite and I make fun of everyone  
I'm immature but I will stay this way forever

Until the day I die I promise I won't change  
so you better give up  
I don't wanna be told to grow up  
and I don't wanna change  
I just wanna have fun  
I don't wanna be told to grow up  
and I don't wanna change  
so you better give up  
cause I'm not gonna change  
I don't wanna grow up

I don't wanna be told to grow up (grow up,  
grow up)  
I don't wanna be told to grow up (grow up,  
grow up)

I don't wanna be told to grow up  
and I don't wanna change  
I just wanna have fun

I don't wanna be told to grow up
and I don't wanna change
so you better give up
I don't wanna be told to grow up
and I don't wanna change
I just wanna have fun
I don't wanna be told to grow up
And I don't want to change
So you better give up
No I don't want to change
So you better give up
Cause I'm not gonna change
I don't wanna grow up!

Texto 19: Letra da música *Grow up* – gravação de Simple Plan, 2002

Nitidamente, a letra da música “*Grow up*”, gravada pela banda Simple Plan, demonstra a perspectiva do enunciador sobre seu jeito de ser, ele não quer mudar, não quer ser “mandado”. Biber (1988, p.111) afirma que a argumentação pode ser explícita marcando a persuasão do próprio falante (seu ponto de vista), ou quando tenta persuadir o ouvinte. No caso dessa música, o autor tanto expõe seu ponto de vista quanto tenta persuadir seu ouvinte a “deixa-lo em paz”.

A letra da música apresenta predominância de verbos no infinitivo (“(to) be”; “to go”; “grow”; “give up”; “have”, etc.), duas ocorrências de modais (“will”; “could”) e três gerúndios (“rocking”; “looking” e “gonna”)<sup>157</sup>. Essas características refletem a função comunicativa de argumentação que o autor deseja expressar. Isto é, ele usa os verbos no infinitivo como elementos de integração entre outros elementos verbais no texto, da mesma forma que faz com os gerúndios, revelando uma unidade argumentativa de sua orientação para não ‘mudar’, culminando na argumentação persuasiva “*you better give up*” (em português, “é melhor você desistir”). Ao mesmo tempo, a letra da música apresenta baixa ocorrência de substantivos, pois o autor não está fornecendo informações específicas sobre pessoas ou objetos, mas preocupa-se com suas motivações e ações.

Já a banda Aerosmith tem mais letras com conteúdo menos argumentativo, desenvolvendo mais o conteúdo informativo de sua mensagem. Vejamos a letra de música com escore mais baixo nesta dimensão: “*Shame on you*”, de 1985.

<sup>157</sup> Nossa interpretação segue a etiquetagem automática do *Tree-tagger*, que considera *gonna* por sua característica básica: *going to*.

Somebody tryin' to take my soul  
 Nobody gotta hear my rock n' roll  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you babe  
  
 No leader, steady as a ghost  
 Tooth sweeter, powdering her nose  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you babe  
  
 A-smith, old hank pank  
 Pulling down your knickers  
 Ah, who you gonna spank  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you babe  
  
 Joe Perry, Oh, Mr Style  
 Some kind of killer bee  
 Buzz a while  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame, shame on you babe  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you babe  
  
 Bogus honey, square dance close  
 Grab your partner, bring it home 'n  
 dosey, dosey  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you baby  
  
 Seemin' jazzy bein' shipped in  
 Wilds of the Stanley  
 Try to given' in  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you baby  
  
 Seven or eleven  
 Go to roll the dice  
 Am I in heaven  
 Or am I in Miami Vice  
  
 Shame, shame, shame on you babe  
 Shame, shame, shame on you baby

Texto 20: Letra da música *Shame on you* – gravação de Aerosmith, 1985

A letra da música é rica em substantivos que é a característica de polo oposto do fator que originou esta dimensão, ou seja, maior frequência de substantivos em oposição a verbos no infinitivo, gerúndio e modais. É preciso lembrar que o fato de uma



característica de um polo ocorrer não implica necessariamente que a característica predominante no polo oposto não ocorra, apenas que a frequência de uma seja mais predominante que de outra.

A presença marcante de substantivos, tais como “*somebody; soul; nobody; rock’n’roll, babe, leader, ghost, knickers, heaven*”, etc., denota a característica muito mais descritiva que persuasiva da letra da música analisada. Em outras palavras, a função comunicativa saliente na letra desta música é a de descrever situações, pessoas e objetos participantes do “momento/vida” do autor.

Quando levamos em consideração a característica **período** de tempo, as letras das músicas do período que vai de 2000 a 2009 apresentam-se mais argumentativas que as da década de 80. Vejamos o quadro descritivo estatístico:

#### Descriptive Statistics

Dependent Variable: Gram\_C\_1

Year	Mean	Std. Deviation	N
0	,4921	2,07156	1200
4	,2296	2,30680	367
5	,1691	2,39430	453
6	,0603	2,21280	1305
Total	,0263	2,14360	6290
9	-,1122	2,09446	1046
7	-,2844	1,98315	1087
8	-,2859	2,04679	832

Quadro 28: Estatística descritiva para o fator Gram\_C1, em relação ao ano

Os números da primeira coluna (*Year*) representam as décadas de gravação das músicas: 4 para os anos de 1940, 5 para os anos de 1950, 6 para os anos de 1960, 7 para 1970, 8 para 1980, 9 para 1990 e 0 para os anos de 2000 a 2009.

As letras das músicas apresentadas anteriormente também se enquadram nesta esfera, todavia, não são as mais representativas, mesmo considerando que a primeira foi gravada em 2002 e a segunda em 1985. Isso acontece porque no agrupamento por década de gravação, as letras de músicas deixam de ser consideradas por artistas e, um artista que não tem a totalidade de suas letras predominantes sob uma perspectiva pode apresentar uma ou outra que tenham escores maiores individualmente. Assim, vemos as mais representativas nas duas polaridades, quando observado o período de gravação.

If I had just one thing to say  
 Before my last breath fades away.  
 Keep your head way up in the clouds  
 And never let them get you down!  
 Never let them get you down!

Keep hoping and dreaming and you will soar!  
 With a little faith and love.  
 You will soar!

If I had just two words to say,  
 To explain my life away.  
 I could say ups and downs,  
 Because, I always kept my defense down.  
 Always kept my defense down!

Keep hoping and dreaming and you will soar!  
 With a little faith and love.  
 You will soar!

I wish I could take you away  
 To a place of mystery  
 We will travel there on sands of time  
 To listen to fairy tales  
 We will listen to fairy tales  
 We will listen, listen, listen

Keep hoping and dreaming and you will soar  
 With a little faith and Love.  
 You will soar!  
 You will soar!  
 You will soar!

Texto 21: Letra da música *You will soar* – gravação de Creed, 2005

A letra da música “*You will soar*”<sup>158</sup> da banda Creed, de 2005, é a mais representativa desta dimensão sob a perspectiva temporal. A letra apresenta muita repetição do modal “*will*” e dos infinitivos, o que representa o discurso argumentativo. Enquanto os infinitivos contribuem para a unidade do texto (“*say*”; “*get*”; “*soar*”; “*listen*”, etc.), os modais “*will*” e “*could*” estabelecem as consequências das ações do ouvinte. Em suma, mesmo apresentando poucos referenciais concretos (“*breath*; *soul*; *faith*; *love*”), o autor pretende convencer seu ouvinte de que, se ele continuar a ser um sonhador ele vai sofrer.

<sup>158</sup> Na realidade, a música foi gravada pela banda Alter Bridge, que foi formada com todos os integrantes da banda Creed, menos o vocalista, durante um hiato entre as gravações da banda original, e, por isso, foi categorizada por nós como parte do arquivo de músicas do Creed.

Já na letra da música “*Radio Waves*”<sup>159</sup>, gravada em 1987 por Roger Waters, do Pink Floyd, o autor apresenta suas perspectivas de forma descritiva. Isto é, fala sobre a atmosfera, a camada de ozônio, a visão de “*Billy*”, etc., tudo nas “ondas do rádio”, ou seja, apresenta diversos elementos marcadores de situação e informação (substantivos), sem nenhuma tentativa direta de expor sua opinião, ou de argumentar, ou persuadir o ouvinte de forma direta (baixa frequência de gerúndios, nenhum infinitivo ou modal).

Radio waves, Radio waves,  
 He hears radio waves, Radio waves,  
 The atmosphere is thin and cold  
 The yellow sun is getting old  
 The ozone overflows with radio waves  
 AM, FM whether and news  
 Our leaders have a frank exchange of views  
 Are you confused, radio waves.

Radio waves, radio waves  
 AM radio waves, FM radio waves  
 Radio waves, mind-numbing radio waves  
 Fish-stunning radio waves  
 Radio waves.

Magic Billy in his wheel chair  
 Is picking up all this stuff in the air  
 Billy is face to face with outer space  
 Messages from distant stars  
 The local police calling all cars, radio waves

Hear them radio waves, radio waves  
 Jesus saves radio, radio waves  
 Radio waves, AM radio waves, FM radio waves  
 All them radio waves

Radio waves, radio waves, he hears radio waves  
 Radio waves, radio waves, hopeful radio waves,  
 dopeful radio waves  
 Radio waves, Russian radio waves, Prussian radio waves  
 Eastern radio waves, western radio waves  
 Testing radio waves, one, two. One, two.  
 Radio waves. Getting through to you  
 Morse code radio waves, Tobacco road radio waves  
 South to Paloma radio waves, Oklaholma City radio  
 waves  
 Sitting pretty radio waves, nitty-gritty radio waves  
 Radio waves

Texto 22: Letra da música *Radio Waves* – gravação de Pink Floyd/Roger Waters, 1987

No cômputo **estilístico** musical nesta dimensão, o *Punk Pop* se opõe ao *Rock surfista*, como demonstra o resultado da análise descritiva reproduzido a seguir:

<sup>159</sup> Esta música faz parte da carreira solo do cantor Roger Waters, mas foi agrupada em nosso corpus como gravação de sua banda original, Pink Floyd.

### Descriptive Statistics

Dependent Variable: Gram\_C\_1

Style	Mean	Std. Deviation	N
PUNK POP	1,3914	1,98342	160
POST GRU	,8806	2,00788	248
POP ROCK	,3704	2,28683	449
ROCK & R	,3159	2,06701	305
Pop (voc	,1363	2,34629	1218
ROCK	,1179	2,13780	806
GRUNGE	,0378	2,50634	183
Total	,0263	2,14360	6290
POP	-,0366	1,95136	591
ALTERNAT	-,1361	1,90048	220
HARD ROC	-,1620	1,97495	421
COUNTRY	-,3435	1,88466	645
PROGRESS	-,3896	2,03025	275
PUNK ROC	-,4473	1,96512	159
HEAVY ME	-,4519	1,98236	334
ROCK (su	-,4902	2,06954	276

Quadro 29: Estatística descritiva para o fator Gram\_C1, em relação estilo musical

A canção de *Punk Pop* mais representativa desta dimensão é “*My number One*” da banda Paramore. Como já apresentamos esta canção no item 5.1.7, ao interpretarmos os fatores léxico-gramaticais – que compartilham as características morfossintáticas com os fatores gramaticais –, ilustraremos esta dimensão com a canção “*Swim in Silence*”, gravada pela mesma banda, em 2009, pois é a segunda mais representativa.

```

Meet me here beneath the burning skies
Where the ocean comes and takes us from all of our lies
You never said that you were coming back
I have waited although I have found the place you hide

What keeps you so far away?

We can swim in silence
You can pull me under
I will not come up for anyone
I can slowly sink and
Watch you as you leave
But I will drown until you care
I will drown until you care

I imagine what it must be like
To have everything you need and not be satisfied
Run the water until it burns
And you can't see through the waves that crash into your
prey

What keeps you so far away?

We can swim in silence

```

You can pull me under  
 I will not come up for anyone  
 I can slowly sink and  
 Watch you as you leave  
 But I will drown until you care

Going under, Going under

Getting close to what we cannot recognize  
 Floating face down in the lies  
 Here we are without a trace  
 But the lies we used to blame  
 and you're so far away  
 What keeps you so far away?

We can swim in silence  
 You can pull me under  
 I will not come up for anyone  
 I can slowly sink and  
 Watch you as you leave  
 But I will drown until you care  
 I will drown until you care

Texto 23: Letra da música *Swim in Silence* – gravação de Paramore, 2009

As características argumentativas desta canção residem na proposta daquilo que os participantes (“we”, “nós”) podem fazer e/ou farão, representada pelos modalizadores “can” e “will” e “must”; os imperativos como “meet”, “watch” e “run” acompanham os modais explicitando a informação argumentativa. Em suma, a autora usa essas características linguísticas para tentar argumentar com seu interlocutor sobre o quanto ela “afunda” sem a presença e o cuidado dele.

No polo oposto, apresentamos a letra da música “*Pom Pom Play Girl*”, da banda de rock surfista Beach Boys, que é altamente descritiva, rica em substantivos e ‘palavras sem sentido’:

Rah rah pom pom play girl  
 Pom pom play girl

(Oooooo)

You see her in a short skirt down on the grass  
 She's wonderin' if the team's gonna run kick or  
 pass  
 (Chi chi bah oooooo)

Since she doesn't really know why she's wavin'  
 her hands  
 She's a-thinking Ôbout the boy sittin' up in  
 the stands

Rah rah pom pom play girl  
 Pom pom play girl

(Chi chi bah)  
 Just a little social clique with the other song  
 queens  
 With a freshman or a sophomore she wouldn't be  
 seen  
 (Chi chi bah oooooo)  
 She's a doll and she knows it and she thinks  
 that her face  
 Will get her any body any thing and any place

Rah rah pom pom play girl  
 Pom pom play girl  
 Oooooo

Chi chi bah

(Oooooo)  
 Since she's the head cheerleader she dates the  
 quarterback  
 And her best friend's going out with the full  
 back  
 (Chi chi bah oooooo)  
 She might even run for an office this Spring  
 And that ought to really make her telephone  
 ring

Rah rah pom pom play girl  
 Pom pom play girl

(Chi chi bah pom pom play girl)  
 Wooooo  
 (Chi chi bah pom pom play girl)  
 Wooo  
 (Chi chi bah pom pom play girl)  
 Shake it  
 (Chi chi bah pom pom play girl)  
 Wave those pom poms all around  
 (Chi chi bah pom pom play girl)  
 (Chi chi bah pom pom play girl)  
 (Chi chi bah pom pom play girl)  
 Wow  
 (Chi chi bah pom pom play girl)  
 Wow  
 (Chi chi bah pom pom play girl)

Texto 24: Letra da música *Pom pom play girl* – gravação dos Beach Boys, 1964

Embora a letra da música apresente ocorrências de gerúndio, infinitivo e modal (apenas uma ocorrência de “*will*”), há predominância de referentes informativos/descritivos, como “*girl; skirt; dress; team; place; face; cheerleader; quaterback*”; entre outros. As escolhas gramaticais do autor denotam a função informativa que ele deseja apresentar, ou seja, a letra informa ao interlocutor as características concretas de uma garota do tipo “pom pom”.

A primeira dimensão gramatical é caracterizada pela oposição de discursos argumentativo e informativo. Essa oposição foi exemplificada sob três diferentes perspectivas, ou seja, no que diz respeito aos artistas que gravaram as músicas, ao períodos de tempo quando foram gravadas e ao estilo musical adotado pela mídia.

Pudemos confirmar, portanto, que, no que diz respeito à dimensão ARGUMENTAÇÃO X INFORMAÇÃO, as letras das músicas gravadas pela banda Simple Plan são mais argumentativas que as letras das músicas gravadas pela banda Aerosmith. Essa relação se mantém, quando observamos a década de gravação das músicas. Isto é, músicas gravadas a partir do ano 2000 tendem a ser mais argumentativas e menos informativas que músicas gravadas nos anos 80 e vice-versa. É interessante observar que a banda Aerosmith tem muitas gravações nos anos 80 e que a banda Simple Plan foi montada depois de 2000. Ainda, no que diz respeito à categorização por gêneros/estilos musicais de acordo com a mídia, as músicas do estilo *punk pop* são mais argumentativas e as de *rock surfista* menos argumentativas e mais informativas.

Os artistas, períodos e estilos não mencionados especificamente na análise encontram-se mais próximos ou mais distantes dos polos característicos da dimensão como podemos observar nos quadros de estatística descritiva apresentados nesta seção. As letras de música do estilo *post grunge*, por exemplo, estão mais próximas das de *punk pop* que as de *pop rock*. Isto significa que, considerando o estilo *punk pop* como detentor de mais características argumentativas, quanto mais distante dele, menos os outros estilos representarão a função argumentativa e, conseqüentemente, mais se aproximarão de um discurso com características informativas, como é o caso do *rock surfista*. Vejamos essa distribuição na forma de um gráfico. No eixo horizontal, estão distribuídos os participantes (aqui, os gêneros musicais). No eixo vertical (“y”) estão calculados os escores-médio em relação ao desvio padrão e ao número de textos (letras de música) para cada participante. A altura das barras representa a quantidade de textos e os participantes no eixo horizontal estão ordenados por escore.

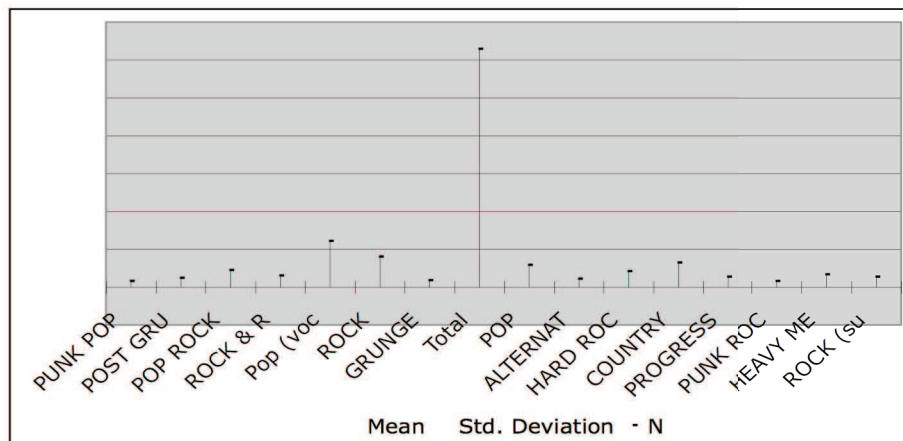


Gráfico 1: Distribuição de estilos musicais na dimensão gramatical 1: Argumentação X Informação

Em relação ao tempo, a distribuição pode ser observada pelo seguinte gráfico:

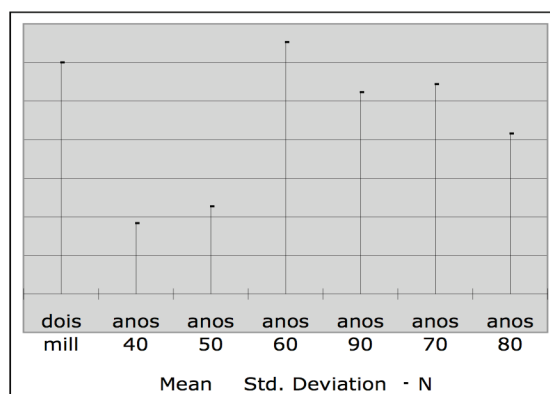


Gráfico 2: Distribuição de período de gravação na dimensão gramatical 1: Argumentação X Informação

Podemos observar no gráfico que as músicas gravadas a partir do ano dois mil são mais representativas desta dimensão em oposição àquelas gravadas nos anos 80. Além disso, as músicas gravadas nos anos 40 são mais parecidas em termos argumentativos com as gravadas a partir de 2000, do que com as gravadas nos anos 60, por exemplo.

Vejamos o que um gráfico nos indica em relação aos artistas:



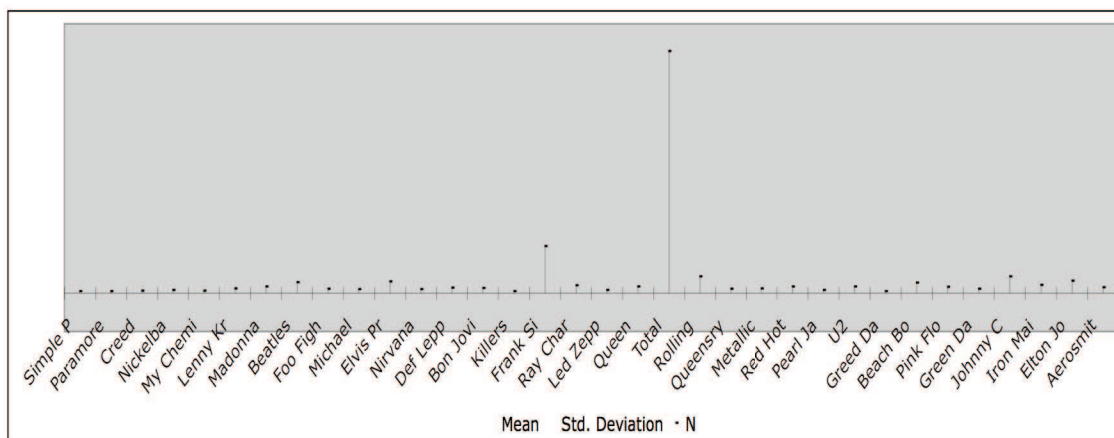


Gráfico 3: Distribuição de artistas na dimensão gramatical 1: Argumentação X Informação

Podemos observar pelo gráfico que artistas como Simple Plan, Paramore, Creed e Nickelback apresentam maior afinidade em termos de características argumentativas. Artistas como Aerosmith, Elton John, Iron Maiden e Johnny Cash também compartilham mais afinidades em termos de discurso informativo em suas letras de música. Bandas como Queen, Rolling Stones e Queensryche apresentam, nas letras de suas músicas, maior equilíbrio entre as duas funções argumentativa e informativa.

Esta análise responde parcialmente às nossas questões de pesquisa que se concentram em desvendar como as dimensões são representadas em relação aos diferentes gêneros musicais, artistas e ao longo do tempo, pois representa apenas uma das dimensões de análise por nós identificadas. Passaremos, portanto, a discutir as dimensões subsequentes.

### 5.2.1.2. Segunda Dimensão Gramatical: Interação versus descrição

A segunda dimensão gramatical é marcada pela alta presença de pronomes pessoais e possessivos, que são usados normalmente para caracterizar um discurso interacional, especialmente quando há predominância de pronomes de primeiras e segundas pessoas, como no caso do nosso corpus. Quando essa característica interacional é bastante frequente, a dimensão é marcada pela baixa frequência de qualificadores. Adjetivos e determinantes são normalmente usados para caracterizar uma informação nominal. Sendo

assim, esta dimensão pode ser nomeada de INTERAÇÃO X DESCRIÇÃO.

Essas três características gramaticais encontram-se na mesma polaridade do fator 1 encontrado na análise de Biber (1988, p. 102, 104-108 e 129-135). Assim como na pesquisa de Biber, as características gramaticais que encontramos aqui aparecem também associadas a outras nas duas polaridades, e tais características levaram o pesquisador a detectar um quadro mais indicativo de um discurso informacional (a extração de Biber agrupou neste polo outras características, como verbos privados<sup>160</sup> no presente, pronomes, enfáticos, indefinidos, entre outros) versus referencial (agrupados também, pela extração fatorial, substantivos, preposições, passivas e participios). Pacheco (1997) também detectou uma dimensão caracterizada pela oposição interacional/informacional, quando contrastou o gênero escrita acadêmica em língua materna e língua estrangeira. O polo interacional foi caracterizado por traços linguísticos típicos, “*como pronomes pessoais de primeira e segunda pessoa*, que indicam interação entre escritor (ex.: ‘eu’) e o leitor (ex.: ‘tu/você’)” (PACHECO DE OLIVEIRA, 1997, p. 138)

Em nossa pesquisa de mestrado (Bértoli-Dutra, 2002) e em nosso estudo piloto, já havíamos apontado para a alta frequência dos pronomes pessoais “*I*” e “*you*”, que, nesta dimensão, caracterizam, sem dúvida, a alta frequência de um discurso interacional, próximo ao conversacional. Veremos como a relação entre interação e descrição se manifesta por meio da observação das características linguísticas nas letras de música mais representativas em cada uma de nossas três perspectivas de análise (artista, década e estilo).

A análise descritiva desta dimensão apontou que a **artista** que mais demonstra um discurso interacional em suas letras de música é a cantora americana Madonna, como podemos observar pelo quadro de estatística descritiva para esta dimensão, reproduzido a seguir:

---

<sup>160</sup> Biber chama de “*private verbs*” (ou verbos psicológicos) verbos que expressam cognição, pensamento, tais como “*antecipar, presumir, acreditar, concluir, demonstrar, decidir, duvidar, estimar, sentir, notar, entender*”, etc. (BIBER, 1988, p. 242)

### Descriptive Statistics

Dependent Variable: Gram\_C\_2

Singer_band	Mean	Std. Deviation	N
Madonna	,8830	1,81979	167
Lenny Kr	,8311	2,61801	125
Nirvana	,7343	2,69188	98
Killers	,6938	1,44964	51
Metallic	,6281	2,46557	119
Michael	,6257	2,03632	99
Beatles	,5960	2,38291	281
Led Zepp	,5732	1,92569	79
Simple P	,4692	1,90211	49
Creed	,4267	1,75679	60
Elvis Pr	,3985	2,31432	305
Paramore	,3149	1,92566	50
Def Lepp	,2833	1,84940	142
Rolling	,2265	2,20927	430
Green Da	,2168	2,31431	110
Bon Jovi	,1441	1,87883	127
Red Hot	,0889	1,95691	169
U2	,0105	2,28673	168
Total	,0012	2,22205	6290
Ray Char	-,0041	2,14540	206
Beach Bo	-,1076	2,82929	276
Queen	-,1091	2,48894	172
My Chemi	-,1155	1,61195	61
Queensry	-,1408	1,75444	112
Frank Si	-,1923	2,34148	1218
Aerosmit	-,1989	1,68702	152
Nickelba	-,2502	2,02882	79
Iron Mai	-,2917	2,02002	215
Greed Da	-,3275	1,76682	49
Elton Jo	-,5059	1,85287	325
Johnny C	-,5155	1,97731	439
Foo Figh	-,6345	1,89347	109
Pearl Ja	-,7088	2,26730	85
Pink Flo	-,7823	2,02706	163

Quadro 30: Estatística descritiva para o fator Gram\_C2, em relação ao artista

A letra da música dessa cantora que mais salienta essas características interacionais é a de “Justify my love”, gravada em 2000. Vejamos:

I wanna kiss you in Paris  
 I wanna hold your hand in Rome  
 I wanna run naked in a rainstorm  
 Make love in a train cross-country  
 You put this in me  
 So now what, so now what?  
  
 Wanting, needing, waiting  
 For you to justify my love  
  
 Hoping, praying  
 For you to justify my love  
  
 I want to know you  
 Not like that  
 I don't wanna be your mother  
 I don't wanna be your sister either  
 I just wanna be your lover  
 I wanna be your baby  
 Kiss me, that's right, kiss me  
  
 Yearning, burning  
 For you to justify my love  
  
 What are you gonna do?  
 What are you gonna do?  
 Talk to me, tell me your dreams  
 Am I in them?  
 Tell me your fears  
 Are you scared?  
 Tell me your stories  
 I'm not afraid of who you are  
 We can fly!  
  
 Poor is the man  
 Whose pleasures depend  
 On the permission of another  
 Love me, that's right, love me  
 I wanna be your baby  
  
 I'm open and ready  
 For you to justify my love  
 To justify my love  
 Wanting, to justify  
 Waiting, to justify my love  
 Praying, to justify  
 To justify my love  
 I'm open, to justify my love

Texto 25: Letra da música *Justify my love* – gravação de Madonna, 2000

Os pronomes pessoais e possessivos de primeira e segunda pessoas são abundantes nessa música (I/we/you/your/me/my). Existem apenas três adjetivos: “*naked*”, “*scared*” e “*poor*”, que aparecem somente uma vez cada. Podemos perceber que na letra a pessoa que fala (canta) parece estar conversando diretamente com o ouvinte. Embora não seja de

forma síncrona, isto é, não exista possibilidade de resposta direta/imediata por parte desse ouvinte – considerando que o mesmo não está presente no momento da concepção da letra –, a mesma pode ser categorizada por suas características interacionais.

Não podemos dizer o mesmo da canção “*Yet another movie*”, gravada em 1987 pela banda inglesa Pink Floyd, que se encontra no polo oposto. Vejamos:

One sound, one single sound  
 One kiss, one single kiss  
 A face outside the window pane  
 However did it come to this?

A man who ran, a child who cried  
 A girl who heard, a voice that lied  
 The sun that burned a fiery red  
 The vision of an empty bed

The use of force, he was so tough  
 She'll soon submit, she's had enough  
 The march of fate, the broken will  
 Someone is lying very still

He has laughed and he has cried  
 He has fought and he has died  
 He's just the same as all the rest  
 He's not the worst, he's not the best

And still this ceaseless murmuring  
 The babbling that I brook  
 The seas of faces, eyes upraised  
 The empty screen, the vacant look

A man in black on a snow white horse,  
 A pointless life has run its course,  
 The red rimmed eyes, the tears still run  
 As he fades into the setting sun

Texto 26: Letra da música *Yet another movie* – gravação de Pink Floyd, 1987

A música apresenta grande quantidade de qualificadores, especialmente adjetivos como “*single*”, “*fiery*”, “*empty*”, “*tough*”, “*broken*”, “*worst*”, “*best*”, “*vacant*”, “*black*”, “*white*”, etc., em oposição à baixa ocorrência, especialmente de primeira e segunda pessoas pronominal (apenas “*it*”, “*he*”, “*she*”, “*its*”) que caracterizam interação. Portanto, a letra dessa música representa muito mais um discurso descritivo que interacional. A letra da música parece apresentar uma descrição de vida e morte de um homem comum que, podemos nos arriscar a dizer (pois a música é bastante

metafórica<sup>161</sup>), aparentemente deixou a casa, o filho.

Quando observamos esta dimensão, considerando o **período** de tempo, as músicas gravadas na década de 40 apresentam mais interação que descrição de que em relação às gravadas nos anos 70, como demonstrado pelo quadro abaixo:

**Descriptive Statistics**  
Dependent Variable: Gram\_C\_2

Year	Mean	Std. Deviation	N
4	,1024	2,36447	367
0	,1019	1,95714	1200
9	,0400	2,21362	1046
5	,0399	2,39720	453
6	,0182	2,42030	1305
Total	,0012	2,22205	6290
8	-,0686	2,15022	832
7	-,1644	2,18018	1087

Quadro 31: Estatística descritiva para o fator Gram\_C2, em relação ao tempo

Frank Sinatra representa o período de 1940 para esta dimensão de análise com a música “*Oh, Bess*”, de 1947, exposta a seguir:

```
Oh, Bess, oh, where's my Bess?
Won't somebody tell me where?
I ain't care what she say, I ain't care what she done
Won't somebody tell me where's my Bess? Bess
My Bess, I want her now
Without her I can't go on
I counted the days that I was gone
Till I got home to see her face
Won't somebody tell me where's my Bess?
I want her so, my gal, my Bess
Where is she?
Oh, Lord, in your big heaven
Please show me where I must go
Oh, give me the strength, show me the way
Tell me the truth, where is she?
Where is my gal?
Where's my Bess?
```

Texto 27: Letra da música *Oh Bess* – gravação de Frank Sinatra, 1940

A letra da música contém apenas uma ocorrência de adjetivo (*big*) e poucos determinantes (*the*), o que simboliza o baixo nível descritivo, sobreposto pelas características interacionais que estão marcadas pela presença frequente de pronomes (*I, she, my, her*). Aqui a interação ocorre de maneira diferente da interação presente na

<sup>161</sup> É preciso observar que, neste momento de análise, estamos investigando apenas os elementos que denotam a dimensão de análise e não outras formas sintáticas ou semânticas.

música “*Justify my love*”, que representou interação entre primeira e segunda pessoas, pois em “*Oh Bess*”, o enunciador usa mais frequentemente pronomes que se referem à terceira pessoa, “Bess”, no caso. Por outro lado, apresenta outros referenciais que indicam que está ‘pedindo a ajuda de alguém’ para encontrá-la (“*won’t somebody tell me*”, em português, ‘ninguém vai me dizer’).

Os anos entre 1970 e 1979 caracterizam-se nesta dimensão por músicas que apresentam mais aspectos descritivos e menos interação, como é o caso de “*Nobody wins*”, também gravada por Sinatra, mas em 1973.

```

Anymore it doesn't matter who's right or wrong.
We've been injurin' each other much too long.
And it's too late to try to save what might have been.
It's over; nobody wins.
Make believin' in forever is just a lie.
And it seems a little sadder each time we try.
'Cause it's a shame to make the same mistake again and
again.
It's over; nobody wins.
We've gone too far, too long, too far apart.
The lovin' was easy; it's the livin' that's hard.
For there's no need to stay to see the way it ends.
It's over; nobody wins.
We've gone too far, too long, too far apart.
The lovin' was easy; it's the livin' that's hard.
For there's no need to stay to see the way that it
ends.
It's over; nobody wins.
It's over; nobody wins.

```

Texto 28: Letra da música *Nobody Wins* – gravação de Frank Sinatra, 1973

A letra de música apresenta menor ocorrência pronominal em contraste à ampla ocorrência de qualificadores como adjetivos e advérbios, como “*right, wrong, late, forever, sadder, same, again, too, far, easy,*” etc. Podemos perceber, portanto, que, embora exista uma característica interacional presente na letra, as descrições sobre a situação entre “eles” agora sobrepõem-se aos traços interacionais. Desse modo, podemos classificar a música como mais descritiva que interacional.

Os **estilos** musicais que ocupam as posições extremas nesta dimensão são o *rock’n’roll* e o *rock progressivo*.

**Descriptive Statistics**  
Dependent Variable: Gram\_C\_2

Style	Mean	Std. Deviation	N
ROCK & R	,3985	2,31432	305
POP ROCK	,3769	2,36191	449
ROCK	,2827	2,32843	806
ALTERNAT	,2291	1,86627	220
PUNK POP	,1981	1,81113	160
POP	,0761	1,98176	591
HARD ROC	,0672	1,80933	421
GRUNGE	,0640	2,59913	183
PUNK ROC	,0490	2,16953	159
HEAVY ME	,0360	2,22958	334
Total	,0012	2,22205	6290
ROCK (su	-,1076	2,82929	276
Pop (voc	-,1923	2,34148	1218
POST GRU	-,2553	1,94451	248
COUNTRY	-,3522	2,04481	645
PROGRESS	-,5210	1,94321	275

Quadro 32: Estatística descritiva para o fator Gram\_C2, em relação ao gênero musical

O representante do *rock'n'roll* em nosso corpus é o cantor Elvis Presley, cuja gravação de “*It's your baby you rock it*”, de 1970 é a mais representativa e está reproduzida a seguir:

```

You offered me a penny for my thoughts
And I told you then that woman won't stay caught
But you turned and loved her anyway
She broke your heart and all I've got to say

It's your baby, you rock it
It's your heartache, you bought it
You made the bed your sleeping in
And I'm tired of hearing about it friend
It's your baby you, rock it

Well you cried upon my shoulder like a baby
I'm sorry 'bout your troubles and your lady
But she done you like she done me
And I've used up all my sympathy
It's your baby you, rock it

It's your baby, you rock it
It's your heartache, you bought it
You made the bed your sleeping in
And I'm tired of hearing about it friend
It's your baby you, rock it

```

Texto 29: Letra da música *It's your baby, you rock it* – gravação de Elvis Presley, 1970

O cantor parece estar conversando com o ouvinte (“*you offered me*”; “*I told*”



*you*”<sup>162</sup>), ele quer que o ouvinte pare de choramingar por seus problemas, porque ele está cansado de ouvi-los e expõe sua opinião diretamente ao interlocutor.

Por outro lado, a letra da música “*On the turning away*”, gravada pelos britânicos do Pink Floyd, em 1987, apresenta vários elementos qualificadores, como “*pale*”, “*downtrodden*”, “*that*”, “*a*”, “*the*”, etc., em contraste à pouca ocorrência pronominal, como podemos observar a seguir:

On the turning away  
 From the pale and downtrodden  
 And the words they say  
 Which we won't understand  
 "Don't accept that what's happening  
 Is just a case of others' suffering  
 Or you'll find that you're joining in  
 The turning away"

It's a sin that somehow  
 Light is changing to shadow  
 And casting it's shroud  
 Over all we have known  
 Unaware how the ranks have grown  
 Driven on by a heart of stone  
 We could find that we're all alone  
 In the dream of the proud

On the wings of the night  
 As the daytime is stirring  
 Where the speechless unite  
 In a silent accord  
 Using words you will find are strange  
 And mesmerized as they light the flame  
 Feel the new wind of change  
 On the wings of the night

No more turning away  
 From the weak and the weary  
 No more turning away  
 From the coldness inside  
 Just a world that we all must share  
 It's not enough just to stand and stare  
 Is it only a dream that there'll be  
 No more turning away?

Texto 30: Letra da música *On the turning away* – gravação de Pink Floyd, 1987

A música é muito mais descritiva que interacional. Há poucos marcadores pronominais de interação como “*we*”, “*you*” e “*they*”, ainda assim, com baixa ocorrência. O compositor pretende que o ouvinte perceba as características dos

<sup>162</sup> Em português: “você me ofereceu”, “eu lhe disse”.

problemas que o (os) rodeiam.

A segunda dimensão gramatical contrapõe elementos linguísticos interacionais e descritivos em letras de música marcados por pronomes pessoais/possessivos em oposição a elementos qualificadores. Percebemos que, no que diz respeito às funções INTERAÇÃO X DESCRIÇÃO, o conjunto das letras de música gravadas pela cantora americana Madonna apresenta muito mais características interacionais que o conjunto de letras de músicas gravadas pelos outros artistas do corpus. Isso não significa dizer que outro artista não tenha uma letra de música que, individualmente, apresente mais características interacionais que a letra de música que analisamos (*"It's your baby you rock it"*), pois, esta análise privilegia o cantor ou banda.

Como dissemos desde o início deste trabalho, a análise multidimensional prevê um arco dimensional, isto é, textos que tendem mais para uma característica e menos para outra, mas que não necessariamente excluem uma ou outra. Vejamos o gráfico dessa variação para esta dimensão quanto ao artista:

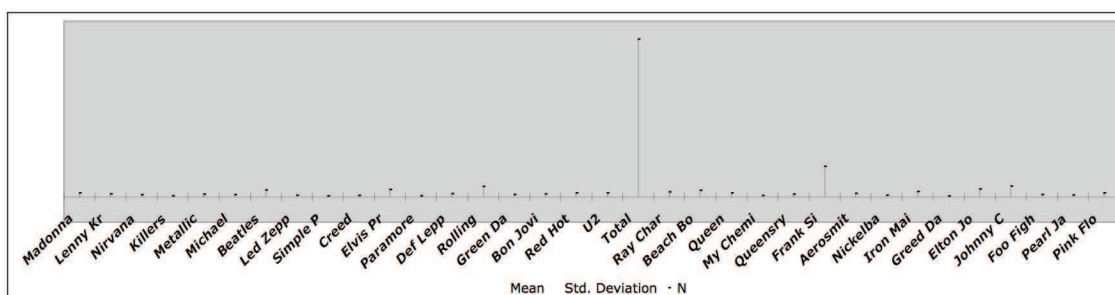


Gráfico 4: Distribuição de artistas na dimensão gramatical 2: Interação X Descrição

Podemos observar pelo gráfico que os artistas que apresentam mais características interacionais em suas letras são os americanos Madonna, Lenny Kravitz, Nirvana, Killers e Michael Jackson. Já artistas como Pink Floyd, Pearl Jam, Foo Fighters e Johnny Cash têm em suas letras características mais descritivas, o que não significa que não possuam letras interacionais. U2 e Ray Charles encontram-se em posição mais central no quadro, o que significa que apresentam letras mais equilibradas quanto à interação e à descrição.

Ao observarmos essas mesmas características em relação ao período em que as músicas foram gravadas, verificamos que as músicas dos anos 40 se apresentam mais

interacionais que as dos anos 70. Vejamos o gráfico:

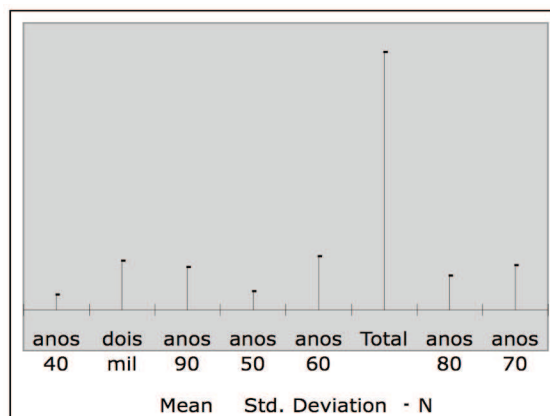


Gráfico 5: Distribuição de período de gravação na dimensão gramatical 2: Interação X Descrição

Como vemos, músicas compostas na década de quarenta e dois mil são mais semelhantes em termos interacionais, e as músicas compostas nas décadas de 80 e 70 apresentam mais elementos descritivos. Verifiquemos como se dá tal variação em termos de gêneros musicais:

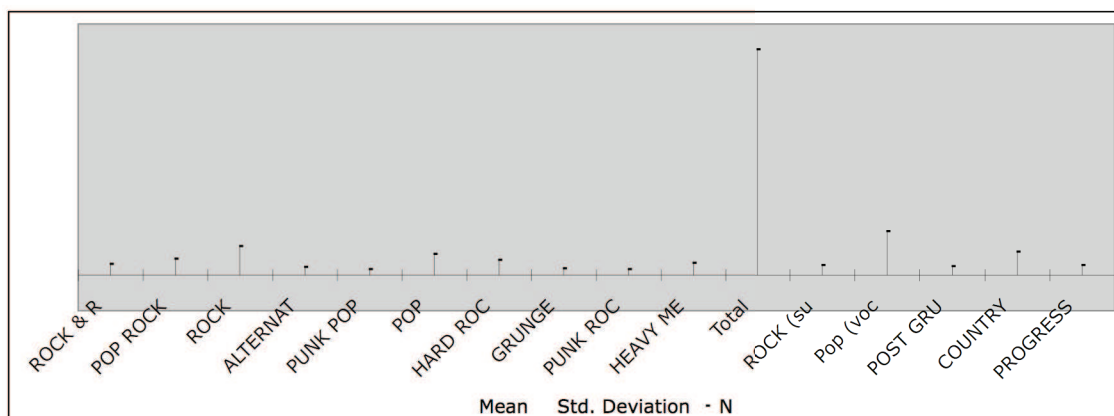


Gráfico 6: Distribuição de estilos musicais na dimensão gramatical 2: Interação X Descrição

O gráfico demonstra que, em termos dos elementos linguísticos que denotam a função interacional do discurso, os estilos *rock'n'roll*, *pop rock* e *rock* apresentam-se mais interacionais que os estilos *progressivo* e *country*. Já o *heavy metal* e o *rock surfista* apresentam letras de música com maior equilíbrio entre esses dois pólos.

Os três gráficos que acabamos de apresentar mostraram a distribuição dos artistas, do tempo e dos estilos musicais nesta dimensão. Ao olharmos para o elemento divisor dos participantes dos três gráficos, o “total”, notamos que há uma tendência maior para

a interação nas letras de música de maneira geral, pois há maior concentração de artistas, tempo ou gênero no polo positivo dos gráficos. Temos dezoito artistas, em oposição a quinze; cinco épocas, em relação a duas e dez estilos musicais em relação a cinco. Essa constatação evidencia que, de maneira geral, independente de se considerarem os aspectos de tempo, de estilo ou as diferenças entre os artistas, as letras de música apresentam um tipo de discurso que tende para o interacional.

### **5.2.1.3. Terceira Dimensão Gramatical: Narrativa no passado versus referencial imediato**

A terceira dimensão gramatical é marcada linguisticamente pela oposição de tempo verbal entre passado (+)<sup>163</sup> e presente (-). Construções verbais de passado (particípio, inclusive), normalmente constituem processos narrativos sequenciais no passado, característicos de um discurso narrativo, usado para criar imagens descritivas com foco na sequência temporal. Construções verbais no presente (gerúndio inclusive) são usadas quando lidamos com questões mais imediatas, muitas vezes associadas a discursos não planejados. Portanto, esta dimensão é caracterizada tipicamente por relatos de eventos no passado ou que lidam com questões imediatas.

Destacamos que esta dimensão apresenta os tempos verbais na mesma oposição oposta detectada por Biber (1988, p. 102, 108-109, 135-142) na Dimensão 2, que o pesquisador chamou de discurso narrativo versus não-narrativo. Contudo, no fator 2 da pesquisa de Biber, embora predominassem variáveis de passado e presente nos polos opostos, a extração também agrupou outras características linguísticas (cf. quadro 2), o que não aconteceu em nossa extração, a qual apontou apenas para as diferenças de tempo. A diferença de extração fatorial se dá por dois motivos, primeiramente pela seleção de variáveis e, em segundo lugar, pela tipologia dos textos estudados. A pesquisa de Kauffman (2005), por exemplo, não apresenta oposição entre passado e presente no mesmo fator e Pacheco (1997) não considerou a diferença temporal em sua análise, mas outros aspectos da manifestação verbal.

---

<sup>163</sup> O sinal de mais entre parênteses (+) simboliza a polaridade positiva e o sinal de menos (-) simboliza a polaridade negativa na extração fatorial.

A análise estatística descritiva da terceira dimensão gramatical em nosso estudo oferece os artistas, as décadas e os estilos mais representativos, conforme podem ser observados nos quadros dispostos nesta seção. O primeiro, expõe a análise em relação aos **artistas**.

**Descriptive Statistics**  
Dependent Variable: Gram\_C\_3

Singer_band	Mean	Std. Deviation	N
Johnny C	1,8469	1,96252	439
Elton Jo	,5786	1,69700	325
Pink Flo	,3348	2,08010	163
My Chemi	,2299	1,30690	61
Aerosmit	,2198	1,80690	152
Killers	,1818	1,55121	51
Nickelba	,1329	1,70817	79
Elvis Pr	,0506	1,98157	305
Frank Si	,0374	2,13395	1218
Paramore	,0060	1,66426	50
Nirvana	,0054	2,27556	98
Total	,0023	1,99628	6290
Ray Char	-,0326	1,96526	206
Bon Jovi	-,0750	1,60583	127
Foo Figh	-,0954	2,16420	109
Rolling	-,1218	1,98312	430
Beach Bo	-,1711	1,96687	276
Red Hot	-,1988	1,71974	169
Green Da	-,2307	1,76518	110
Lenny Kr	-,2438	1,96488	125
Def Lepp	-,2774	1,64301	142
Creed	-,2930	1,68457	60
Led Zepp	-,3331	1,33677	79
Michael	-,3483	1,63700	99
Simple P	-,3923	1,86788	49
Pearl Ja	-,4084	1,77921	85
Queen	-,4591	1,72301	172
Iron Mai	-,4750	1,77619	215
Greed Da	-,5124	1,66216	49
U2	-,5129	1,84427	168
Madonna	-,5789	1,80456	167
Beatles	-,6614	2,00871	281
Queensry	-,8285	1,48978	112
Metallic	-,9488	1,62635	119

Quadro 33: Estatística descritiva para o fator Gram\_C3, em relação a artistas

Podemos ver que o cantor mais representativo desta categoria é Johnny Cash e a banda menos representativa é Metallica. Isso significa que, ao se observar a totalidade das letras das músicas gravadas por aquele artista, haverá uma predominância de

narrativas no passado, em oposição à totalidade das letras das músicas gravadas por este grupo que apresenta predominância de tempo presente. Na realidade, as músicas de Johnny Cash são as mais representativas nas três esferas de análise, como podemos observar nos quadros que seguem, primeiramente, em relação ao **período** de gravação.

Dependent Variable:Gram\_C\_3

Year	Mean	Std. Deviation	N
7	,3986	1,98874	1087
6	,1499	2,22413	1305
Total	,0023	1,99628	6290
5	-,0852	1,95464	453
9	-,1176	1,89882	1046
4	-,1498	2,19519	367
0	-,1907	1,73610	1200
8	-,2032	1,94342	832

Quadro 34: Estatística descritiva para o fator Gram\_C3, em relação ao período de gravação

O ano de 1970 é o mais representativo, com músicas de Johnny Cash carregando os valores fatoriais mais altos, que representam narrativas no passado. Os anos de 1980 são caracterizados por preocupações mais momentâneas, sendo representados também pela banda Metallica, porém, os fatores mais altos ficaram com a banda britânica U2.

Johnny Cash também representa o **estilo** mais característico desta dimensão, que é o *country*.

Descriptive Statistics

Dependent Variable:Gram\_C\_3

Style	Mean	Std.Deviation	N
COUNTRY	1,2466	2,14895	645
POP	,0963	1,79809	591
ROCK & R	,0506	1,98157	305
Pop (voc	,0374	2,13395	1218
Total	,0023	1,99628	6290
PUNK POP	-,0306	1,61807	160
HARD ROC	-,0369	1,70221	421
POST GRU	-,0704	1,91644	248
ALTERNAT	-,1106	1,68643	220
PROGRESS	-,1390	1,94555	275
ROCK (su	-,1711	1,96687	276
GRUNGE	-,1868	2,06486	183
ROCK	-,2334	1,87426	806
PUNK ROC	-,3175	1,73375	159
POP ROCK	-,6058	1,94805	449
HEAVY ME	-,6438	1,73677	334

Quadro 35: Estatística descritiva para o fator Gram\_C3, em relação ao estilo musical

O *heavy metal* predomina o polo oposto ao *country*, no que diz respeito a discurso narrativo de passado em oposição a referencial imediato. A banda mais representativa aqui é o conjunto inglês *Iron Maiden*, seguido pelos americanos do *Metallica*.

Os quadros apontam a maior representatividade para Johnny Cash, vejamos, portanto, duas de suas letras *country*, sendo uma pertencente à década de 1970. Primeiramente, “*You beat all I ever saw*”, gravada em 1969, seguida por “*I saw a man*”, de 1973.

<b>you beat all I ever saw</b>
<p>I walked though ev'ry town saw fortunes lost and found            And when your trail failed I walked holes in both my soles            But I don't expect you back you're somewhere makin' tracks            I crossed your burning bridges and walked through miles of sand            Met the lawless and the law but you beat all I ever saw</p> <p>I dreamed a million miles about your eyes and smiles            I tried to love the best and to turn from all the rest            But I'd scan the skies for you and I only saw your hue            They threw away your mold you were made of frozen gold            And your heart would never thaw you beat all I ever saw</p>
<b>I saw a man</b>
<p>Last night I dreamed an angel came he took my hand he called my name            He bid me look the other way I saw a man I heard him say</p> <p>He said if I be lifted up I'll draw all men to me            He turned and then I saw those nail scared hands that bled for me            I touched the hem of his garment that fell round him there            My life my heart I gave my soul was in his care</p> <p>When I awoke my heart beat so and in the dark I saw a glow            This was no dream he turned my way and again I heard Lord Jesus say</p> <p>He said if I be lifted up I'll draw all men to me            He turned and then I saw those nail scared hands that bled for me            I touched the hem of his garment that fell round him there            My life my heart I gave my soul was in his care</p>

Texto 31: Letras das música *You beat all I ever saw* e *I saw a man*—

gravação de Johnny Cash, 1969 e 1973

As duas letras claramente apresentam um perfil narrativo sequencial de passado. Na primeira, o autor conta tudo que já fez na vida e mesmo assim só não conseguiu “seduzir”, “amolecer” (“*thaw*”) o coração da pessoa amada. Os referenciais de tempo estão marcados pelos verbos no passado, como “*walked*”, “*saw*”, “*failed*”, “*crossed*”, “*met*”, “*dreamed*”, etc. A letra apresenta verbos em outros tempos verbais também,

como “*don’t expect*”, “*beat*” (presente simples); “*you’re making*” (presente contínuo/gerúndio); “*to love*”, “*to turn*” (infinitivos); “*would scan*” (modal), contudo o tempo predominante é o passado. O autor narra tudo o que viu e fez para dizer que nada se compara ao seu interlocutor. Na segunda, o autor relata um sonho no qual teria encontrado um anjo e Jesus, a quem teria oferecido sua alma. Sua narrativa está marcada pelos verbos no passado como “*dreamed*”, “*came*”, “*called*”, “*bid*”, “*saw*”, “*heard*”, “*said*”, “*lifted*”, “*turned*”, etc., em contraste à baixa frequência de verbos em outros tempos como “*say*” e “*I’ll draw*”, com duas ocorrências cada.

Johnny Cash representa esta dimensão, quando consideramos narrativas no passado, nas três esferas de análise, por artista, por período de gravação – pois a letra da música que mais representa a dimensão nos anos 70 é a exposta anteriormente – e pelo estilo *country*. O polo oposto ao passado, quando observamos as características linguísticas de referencial imediato, verbos no presente; na esfera **artista** é representada principalmente pela banda Metallica, como vimos no quadro 33. A seguir reproduzimos a letra desta banda que é a mais representativa:

No life till leather
We are gonna kick some ass tonight
We got the metal madness
When our fans start screaming
It's right well alright
When we start to rock
We never want to stop again
Hit the lights
Hit the lights
Hit the lights
You know our fans are insane
We are gonna blow this place away
with volume higher
Than anything today the only way
When we start to rock
We never want to stop again
Hit the lights
Hit the lights
Hit the lights
With all our screaming
We are gonna rip right through your brain
We got the lethal power
It is causing you sweet pain Oh sweet pain
When we start to rock
We never want to stop again
Hit the lights
Hit the lights
Hit the lights

Texto 32: Letra da música *Hit the lights* – gravação de Metallica, 1983



Em “*Hit the lights*” o autor relata o que costuma fazer depois que começa a tocar e o que vai fazer junto aos fãs enquanto toca. Não há nenhum verbo no passado, a não ser “*got*”, que carrega sentido de posse, com referencial temporal imediato. Sendo assim, a letra encontra-se no extremo oposto à “*You beat all I ever saw*”, analisada anteriormente.

A seguir, apresentamos a letra da música mais representativa do polo ‘referencial imediato’, quando observada a esfera **período** de gravação.

```
Oh to sing this song makes me angry
I'm not angry with you.
Is that all?
Is that all?
Is that all?

Oh to sing this song makes me happy
I'm not happy with you.
Oh to sing this song makes me dance.
Is that all?
Is that all?
Is that all?

Is that all?
Is that all?
Is that all?
Is that all you want from me?
```

Texto 33: Letra da música *Is that all?* – gravação do U2, 1980

Nessa música, o autor questiona o ouvinte sobre suas intenções imediatas em relação à mesma música e apresenta os sentimentos do compositor. Na realidade, a letra é uma repetição de um questionamento situado em tempo presente (“Isso é tudo?”), sem nenhuma referência temporal ao passado.

A letra da música apresentada a seguir é a mais representativa para o polo negativo, quando observado o estilo musical, ou seja, *country* é o estilo mais representativo do polo positivo, enquanto o *Heavy Metal* representa o polo oposto. Vejamos a letra de “*Die with your boots on*” da banda Iron Maiden, caracterizada por esse estilo.

```
Another Prophet of Disaster
Who says the ship is lost,
Another Prophet of Disaster
Leaving you to count the cost.
Taunting us with Visions,
Afflicting us with fear,
Predicting War for millions,
```

In the hope that one appears.

No point asking when it is,  
 No point asking who's to go,  
 No point asking what's the game,  
 No point asking who's to blame,  
 'Cos if you're gonna die, if you're gonna die,  
 'Cos if you're gonna die, if you're gonna die,  
 If you're gonna die, if you're gonna die,  
 If you're gonna die, if you're gonna die,

If you're gonna die, die with your boots on,  
 If you're gonna try, well stick around,  
 Gonna cry, just move along,  
 If you're gonna die, you're gonna die.

13 the Beast is Rising,  
 The Frenchman did surmise,  
 Through earthquakes and starvation,  
 The Warlord will arise,  
 Terror, Death, Destruction,  
 Pour from the Eastern sands,  
 But the truth of all predictions,  
 Is always in your hands.

You're gonna die, die with your boots on,  
 If you're gonna try, well stick around,  
 Gonna cry, just move along,  
 If you're gonna die, you're gonna die.

They die with their boots on,  
 Yes they die.

The day, they die with their boots on  
 The day we die, we die, we die, we die, we die

No point asking when it is,  
 No point asking who's to go,  
 No point asking what's the game,  
 No point asking who's to blame,  
 'Cos if you're gonna die, if you're gonna die,  
 'Cos if you're gonna die, if you're gonna die,  
 If you're gonna die, if you're gonna die,  
 If you're gonna die, if you're gonna die,

If you're gonna die, die with your boots on,  
 If you're gonna try, well stick around,  
 Gonna cry, just move along,  
 If you're gonna die, you're gonna die.

Die!

Texto 34: Letra da música *Die with your boots on* – gravação do Iron Maiden, 1983

Na letra da música “*Die with your boots on*”, o autor alerta o ouvinte sobre a presença de falsos profetas apocalípticos no presente e aconselha a tomar atitudes. O referencial imediato está marcado pela presença de verbos no presente (*says, is, die,*

*leaving, afflicting, predicting, going, etc.*). Não existem marcadores de passado, apenas indicações sobre consequências ou sugestões para o tempo futuro.

Pudemos confirmar, portanto, que no que diz respeito à dimensão NARRATIVA NO PASSADO X REFERENCIAL IMEDIATO, as letras das músicas gravadas pelo cantor e compositor Johnny Cash referem-se mais ao passado do que as letras das músicas gravadas pela banda Metallica. Essa relação se mantém quando observamos a década de gravação das músicas, ou seja, músicas gravadas nos anos 70 tendem a ser mais enraizadas em narrativas do passado e menos no presente que músicas gravadas nos anos 80 e vice-versa. Já no que diz respeito à categorização por gêneros/estilos musicais de acordo com a mídia, as músicas do estilo *country* apresentam maior preocupação com as “coisas do passado, enquanto às músicas de *heavy metal* estão mais preocupadas com o momento presente. Os outros artistas, estilos e épocas se distribuem entre esses, apresentando-se mais ou menos para cada um dos polos da dimensão. O quadro que segue apresenta essa distribuição em termos de artista.

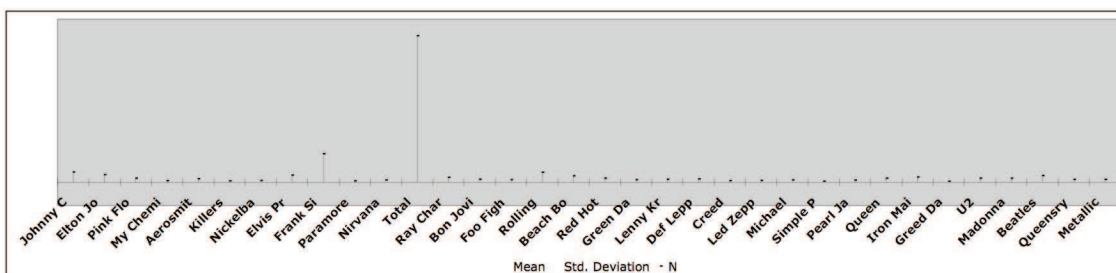


Gráfico 7: Distribuição de artistas na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato

Como podemos ver, em relação ao tempo, artistas como Johnny Cash, Elton John e Pink Floyd, embora representantes de estilos musicais diferentes, têm suas letras mais voltadas para narrativas no passado, enquanto artistas como Metallica, Queensrÿche e Beatles mostram-se mais preocupados com o presente. O quadro também nos mostra que há maior concentração de referência imediata, pois, dos 32 artistas presentes em nosso corpus, 21 pendem para o polo negativo desta dimensão.

Ao observarmos o período de gravação, verificamos a mesma tendência de concentração para ações imediatas, como podemos ver no gráfico que segue.

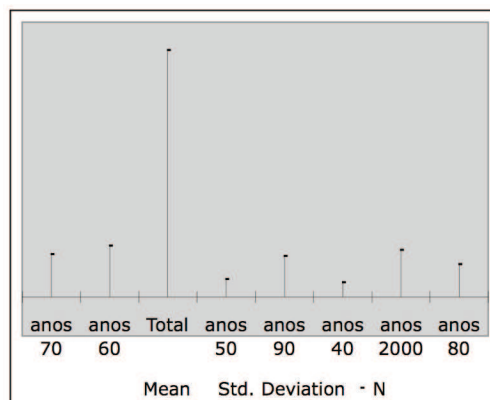


Gráfico 8: Distribuição de período de gravação na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato

As letras das músicas gravadas nos anos 70 e sessenta são as que se destacam com o uso de narrativas no passado de forma superior ao presente. Em todas as outras épocas há predominância de referentes imediatos. Da mesma forma, essa distribuição acontece quando observados os estilos musicais no gráfico seguinte.

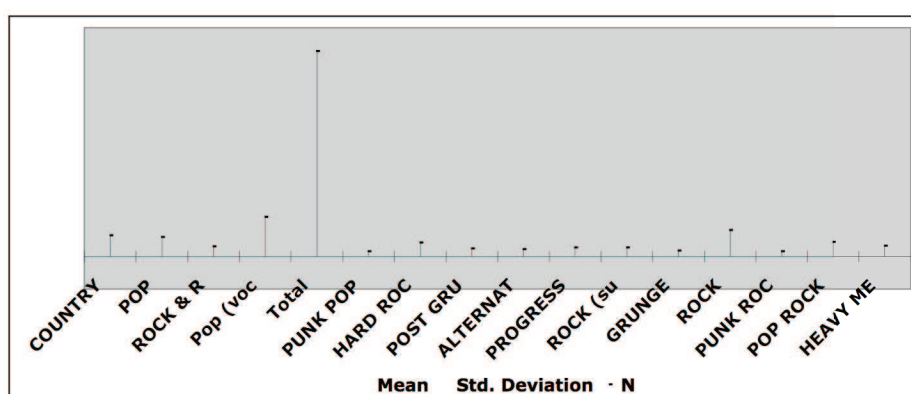


Gráfico 9: Distribuição de estilos musicais na dimensão gramatical 3: Narrativa no Passado X Referencial Imediato

O gráfico mostra que dos quinze estilos diferentes por nós analisados, apenas quatro concentram-se em narrativas no passado: *country*, *pop*, *rock'n'roll* e *pop vocal*. Estilos como *heavy metal* e *pop rock* apresentam maiores semelhanças temporais.

Quando levamos em consideração exclusivamente as características gramaticais presentes em letras de música, artistas, estilos e épocas de gravação, considerados tão distintos por suas características musicais e mercadológicas, compartilham semelhanças linguísticas. As letras das músicas de Frank Sinatra, por exemplo, estão mais próximas das de Elvis Presley (cujo estilo foi amplamente criticado pelo crooner) e da banda de

punk pop deste século, Paramore, do que das letras das músicas gravadas por Ray Charles, por exemplo, quando consideramos os referenciais de tempo. Quando consideramos as características descritivas, as músicas de Frank Sinatra estão mais próximas das gravadas pelas bandas Aerosmith, do estilo *hard rock* e Queensryche, do estilo *progressivo*. Em contrapartida, Sinatra e Ray Charles, que por muitas vezes gravaram versões diferentes para as mesmas músicas, apresentam características linguísticas mais semelhantes quando tem as letras das músicas que gravaram observadas pela função argumentativa.

A análise das dimensões gramaticais para letras de música anglo-americanas podem ser resumidas pelo seguinte quadro:

DIMENSÃO		ARGUMENTAÇÃO X INFORMAÇÃO	INTERAÇÃO X DESCRIÇÃO	NARRATIVA PASSADA X REFERENCIAL IMEDIATO
características		infinitivos; gerúndios; imperativos e modais X substantivos	pronomes pessoais e possessivos X adjetivos e determinantes	verbos no passado X verbos no presente
mais representativos (polo positivo)	artista	Simple Plan	Madonna	Johnny Cash
	período	Anos 2000	Anos 40	Anos 70
	estilo	Punk Pop	Rock'n'roll	Country
menos representativos (polo negativo)	artista	Aerosmith	Pink Floyd	Metallica
	período	Anos 80	Anos 70	Anos 80
	estilo	Rock Surfista	Progressivo	Heavy Metal

Quadro 36: Resumo das dimensões gramaticais

Esta análise fornece as informações para respondermos nossas questões de pesquisa quando o foco é gramatical, a seguir passaremos a observar o foco semântico que representará os assuntos mais comumente tratados pelas letras de música em nosso corpus sob três dimensões, contrastado também, em termos de artistas, período de gravação e estilo musical.

### 5.2.2. Dimensões Semânticas

A análise de dimensões semânticas para letras de música anglo-americanas levou em

conta 8 grandes categorias associadas a movimento, emoção, manifestação musical, tempo, fala, pessoas, objetos e questões sociais.

Conforme apresentado na seção 5.1., o programa estatístico SPSS retornou 3 fatores que indicam o comportamento predominante dessas variáveis em nosso corpus de estudo e serão interpretadas como dimensões de análise para letras de música anglo-americana. A análise fatorial mostrou-nos também que, quando observadas por características semânticas, as letras de música não apresentam oposição de polarização, ou seja, as características mais fortes em cada dimensão encontram-se sempre agrupadas no polo positivo, independente da presença de outras características no polo oposto, como vimos na extração fatorial resumida no quadro 23, seção 5.1.

É importante destacar aqui o caráter inovador desta perspectiva de análise, pois as características lexicais semânticas não foram contempladas por nenhuma das pesquisas por nós observadas. Biber (1988) concentrou sua análise em características linguísticas de cunho morfossintático, embora tenha incluído características que chamou de “classes lexicais”, sendo elas palavras de ligação<sup>164</sup> (“*conjuncts*”), moderadores (“*downtoners*”)<sup>165</sup>, salvaguardas<sup>166</sup> (“*hedges*”), amplificadores (e.g., extremamente, fortemente, etc.), palavras de ênfase (e.g., com certeza), partículas discursivas (e.g., bem, de qualquer forma) e demonstrativos (que por nós foram incluídos na classificação gramatical no grupo dos qualificadores) (BIBER, 1988, 239-241). Kauffmann (2005) observou apenas as características de cunho gramatical. Pacheco de Oliveira (1997) observou além das características morfossintáticas e classes lexicais propostas por Biber, às quais acrescentou expressões que estabelecem conexões entre as partes do texto (como relacionadores extra textuais, sequenciais, marcadores de mudança de tópico), dado o objetivo de sua análise de verificar as redes estruturais dos textos de alunos de graduação. Enfim, nossa pesquisa considera tanto características morfossintáticas, explicadas anteriormente, como inova ao considerar o agrupamento de itens lexicais em escalas semânticas.

---

<sup>164</sup> Cf. tradução de Pacheco de Oliveira (1997, p. 95).

<sup>165</sup> Palavras como “quase, raramente, parcialmente”, etc. (BIBER, 1988, p.240).

<sup>166</sup> Cf. tradução de Kauffmann (2005, p. 23); são expressões como “algo como; quase; do tipo”, etc.

Dessa forma, a primeira dimensão é caracterizada por agrupamentos semânticos que indicam movimento, tempo, fala, pessoas e objetos. A segunda dimensão semântica caracteriza-se pela presença de elementos lexicais que denotam emoção e manifestações sociais simultaneamente. Já a terceira dimensão é caracterizada por elementos lexicais que se referem a manifestações musicais. Essas dimensões estão desvendadas a seguir.

### 5.2.2.1. Primeira Dimensão Semântica: Ações pessoais

A primeira dimensão semântica é determinada pela presença marcante de palavras que podem ser associadas a movimento, fala, pessoas, tempo e objetos. O aparecimento simultâneo desse tipo de palavras em letras de músicas sugere que as letras abordem temáticas que envolvam as ações das pessoas. Ou seja, em termos funcionais, essa dimensão deve marcar relatos das pessoas sobre suas ações e/ou de outras pessoas e as partes envolvidas nessas ações, sejam pessoas ou objetos, localizadas ao longo do tempo.

A tabela estatística descritiva desta dimensão em relação a **artistas** aponta os escores médios de cada um para esta dimensão. Assim, é possível ordená-los de forma que se perceba quais gravaram mais músicas cujas letras são mais representativas nesta dimensão. Vejamos o quadr

**Descriptive Statistics**  
Dependent Variable: Sem\_C\_1

Singer_band	Mean	Std. Deviation	N
Beach Bo	,6620	4,48084	276
Elvis Pr	,5939	3,52811	305
Pink Flo	,5315	4,13522	163
Queen	,3898	3,19305	172
Ray Char	,3475	3,55123	206
Beatles	,3126	3,36933	281
Johnny C	,3106	2,99420	439
Rolling	,3025	2,87101	430
Led Zepp	,2611	2,87128	79
Greed Da	,1747	1,68521	49
Michael	,1065	1,98886	99

Foo Figh	,0611	1,61322	109
Total	,0194	2,69490	6290
Frank Si	-,0339	2,90881	1218
Killers	-,0552	1,53327	51
U2	-,0801	2,10562	168
My Chemi	-,0975	2,05566	61
Simple P	-,1369	1,46265	49
Aerosmit	-,2000	1,78997	152
Elton Jo	-,2571	1,75693	325
Paramore	-,2654	1,80655	50
Pearl Ja	-,2877	1,56778	85
Nickelba	-,3124	1,34154	79
Red Hot	-,3238	1,39153	169
Metallic	-,3492	1,47249	119
Green Da	-,3507	1,42875	110
Madonna	-,3567	1,46613	167
Queensry	-,3883	1,46287	112
Iron Mai	-,3922	1,29875	215
Creed	-,4376	1,39745	60
Lenny Kr	-,4852	1,63488	125
Bon Jovi	-,5045	1,21489	127
Nirvana	-,5985	1,62249	98
Def Lepp	-,6290	1,35347	142

Quadro 37: Estatística descritiva para o fator Sem\_C1, em relação a artistas

O quadro anterior apresenta três colunas, a primeira contém o nome do artista; a segunda, “*Mean*”, apresenta o escore médio de frequência das características por artista (média entre a maior e menor frequência); a terceira apresenta o desvio padrão (*Std. Deviation*), que é uma medida de como a característica se distribui entre os textos; e a última marca o número de textos (letras de música) de cada artista.

Podemos observar que o maior representante da dimensão ações pessoais é a banda americana Beach Boys, e o menor representante é o conjunto inglês Def Leppard.

Vejamos a letra da música “*Morning Christmas*”, gravada em 1964 pelos Beach Boys:

Holy holy
Halo glowing
Candle burning
Christmas evening
Cold wind blowing
Old Man Winter
Inward glowing
Outside snowing
Morning Christmas



Children love to play
Children laugh
Children love to sing
A Christmas song
Holy holy
Halo glowing
Candle burning
Christmas evening

Texto 35: Letra da música *Morning Christmas*– gravação de Beach Boys, 1964

A letra da música apresenta-se marcada pelo evento “Natal”, o que estabelece sua relação com o tempo. Apresenta um relato daquilo que as pessoas gostam de fazer e fazem nesse dia, como se movimentam, por meio de expressões como brincar, cantar, rir (“*play*”, “*laugh*”, “*sing*”). Além disso, descreve ações de objetos envolvidos com o evento, como “vela queimando” (“*candle burning*”), “vento soprando (“*wind blowing*”) e dentro de casa “brilhando” e lá fora “nevando” (“*glowing/snowing*”).

Já a letra da música “*She’s too tough*” da banda Def Leppard (1993) apresenta a descrição de uma moça (“*she*”) por suas características e não por ações, como podemos ver a seguir:

Black stockings and a high-heeled dress
Good-looking but her face is a mess
She's no angel, oh no
Love loaded she's got plenty to spare
She'll pull your trigger if only you dare
She's no angel, oh no
She's too tough
Can't help it, She just can't help it
She's too tough
Can't help it, She just can't help it
Hands off, well it's Catch-22
Y' can't touch if she can't have you
She's no angel, oh no...
Whiplash give you six of the best
Lay back let her mind do the rest
She's no angel, oh no.
She's too tough
Can't help it, She just can't help it
She's too tough
Can't help it, She just can't help it
She's too tough
Can't help it, She just can't help it
She's too tough
Can't help it, She just can't help it

Texto 36: Letra da música *She’s too tough*– gravação de Def Leppard, 1993

A letra da música apresenta uma pessoa (*she*) e suas ‘características’, existem alguns referenciais de objeto, como “*face*” e “*dress*” e algumas ações, como “*pull*” e “*touch*”. O verbo “*help*” não apresenta sentido de movimento, mas de reflexão “não consegue evitar”. A letra dessa música não apresenta referencial temporal, nem de fala. Enfim, para ser característica dessa dimensão, a letra de música deve apresentar associações semânticas que indicam movimento, fala, pessoas, tempo e objetos simultaneamente, como a anterior. A baixa incidência concomitante dessas características torna essa letra a menos representativa desta dimensão, na perspectiva artista.

No que diz respeito ao **tempo**, o SPSS retornou a seguinte análise descritiva:

**Descriptive Statistics**

Dependent Variable: Sem\_C\_1

Year	Mean	Std. Deviation	N
6	,5242	3,76269	1305
7	,4477	3,59569	1087
5	,0517	2,85581	453
Total	,0194	2,69490	6290
8	-,1975	1,78283	832
0	-,3249	1,49089	1200
9	-,3580	1,47596	1046
4	-,3903	1,63412	367

Quadro 38: Estatística descritiva para o fator Sem\_C1, em relação ao período de gravação

A observação dos escores médios por década nos leva a concluir que a década de 1960 teve mais músicas que falavam sobre as ações das pessoas em comparação a outras épocas, especialmente os anos 40. Gravações de Frank Sinatra desses dois períodos ocupam os extremos opostos, especialmente pelas canções “*Mountain Greenery*”, de 1963, e “*Neiani*”, de 1940, que apresentamos na seção 5.1.4. A presença desse artista, nos dois extremos da dimensão, demonstra que gravações feitas por um mesmo artista estão distribuídas de acordo com suas características linguísticas e não por suas características musicais. Ou seja, um determinado cantor ou banda gravou diversas músicas, o fato de uma delas ser a mais representativa em uma categoria de análise não significa que todas as suas outras músicas apresentarão a mesma distribuição de característica. Além disso, quando se observa o conjunto das gravações por artista os

pesos estão agrupados, sendo possível, como já dissemos na análise das dimensões gramaticais, que uma música individual de um outro artista tenha um escore maior que a mais representativa da análise por artista.

Relembrando as letras das músicas mais e menos representativas desta dimensão durante a interpretação dos fatores, na seção 5.1.4 (texto 11 e texto 12), percebemos que a primeira, *Mountain Greenery*, apresenta um relato sobre as ações de pintar, chover, rir, procurar, cozinhar e olhar, de duas pessoas em sua casa. Na segunda, *Neiani*, embora tenhamos a presença de pessoas (*I* e *Neiani*) e uma referência de tempo (*nights*), o objetivo da letra é relatar o momento em que se conheceram e comentar sobre a lembrança “dela”, ou seja, sem representatividade de movimento, fala e objetos.

Outra canção de Sinatra que representa esta dimensão ao longo do tempo é “*Emily*”:

Emily, Emily, Emily
Has the murmuring sound of May
All silver bells, coral shells, carousels
And the laughter of children at play
They say 'Emily, Emily, Emily'
And we fade to a marvelous view
Two lovers alone and out of sight
Seeing images in the firelight
As my eyes visualize a family
They see dreamily, Emily
Two lovers alone and out of sight
Seeing images in the firelight
As my eyes visualize a family
They see dreamily, Emily

Texto 37: Letra da música *Emily*– gravação de Frank Sinatra, 1963

A canção descreve o “falar” das crianças, o “murmurar” do som (fala); descreve os movimentos de rir, esvaecer, ver e visualizar; contém objetos como sinos, corais, conchas; está situada no tempo, no mês de maio, e apresenta referências à pessoas “*they*”, “*children*”, “*Emily*” e “*lovers*”. O aparecimento concomitante dessas características, que exprimem movimento, fala, pessoas, tempo e objetos, claramente representa esta dimensão.

Outro aspecto por nós observado foi a dimensão em relação ao **estilo**. Para essa categoria o SPSS extraiu o seguinte quadro de análise descritiva:

### Descriptive Statistics

Dependent Variable: Sem\_C\_1

Style	Mean	Std. Deviation	N
ROCK (su	,6620	4,48084	276
ROCK & R	,5939	3,52811	305
COUNTRY	,3224	3,17997	645
ROCK	,1949	2,80277	806
POP ROCK	,1657	2,96381	449
PROGRESS	,1569	3,34397	275
Total	,0194	2,69490	6290
Pop (voc	-,0339	2,90881	1218
PUNK POP	-,1620	1,80316	160
POST GRU	-,1785	1,49000	248
PUNK ROC	-,1888	1,52652	159
POP	-,2243	1,72651	591
ALTERNAT	-,2616	1,42656	220
HEAVY ME	-,3769	1,36115	334
HARD ROC	-,4365	1,49795	421
GRUNGE	-,4541	1,60050	183

Quadro 39: Estatística descritiva para o fator Sem\_C1, em relação ao estilo/gênero musical

A tabela de análise descritiva em relação aos estilos musicais aponta que, para o nosso corpus, as canções do estilo “*rock surfista*” mostram mais as ações das pessoas, vinculadas ao tempo, à fala e a objetos, do que as músicas do estilo “*grunge*”. A banda americana “Beach Boys” é a única representante do primeiro estilo em nosso corpus e a letra de sua canção “*Morning Christmas*”, primeira letra de música por nós analisada nesta seção (texto 34), é a mais representativa. Apresentaremos, portanto, a letra de uma música do estilo “*grunge*” que contém pouca representatividade de relato de ações de pessoas situadas no tempo, sendo a menos representativa nesta análise.

<p>Hey, the love of two  A desire, is what's for you  Is it real? A lullaby?  Face to face, with an unnamed mouth  Hey! Hey! Heyyyy! Hey! hey! Heyyyy!  Feel the spirit form, in his eyes was loving for  And his lion's were raging high, hunts himself &amp;  fears his ripe  Hey! Hey! Heyyyy! Hey! Hey! Heyyyy!</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Texto 38: Letra da música *Bambi Slaughter*– gravação do Nirvana, 2004

A canção “*Bambi slaughter*”, da banda Nirvana, foi divulgada em 2004, num álbum em homenagem aos 10 anos da morte de seu compositor, Kurt Cobain, contendo gravações inéditas. A música representa pouca movimentação, não contém referências a

fala ou tempo. As palavras “*himself*” e “*his*” são as únicas características que se encaixam nesta dimensão – podendo marcar referências a pessoas, ainda que, especificamente, refiram-se ao “leão” da música – e, portanto, podemos afirmar que é pouco representativa.

Como vimos, é possível observar canções que se agrupam semanticamente por descreverem ações de pessoas situadas no tempo. Ainda que não existam características semânticas que representem uma polaridade negativa à representada nesta dimensão, as letras de música se distribuem como mais ou menos representativas de acordo com perspectiva de análise. Vejamos como se dá esta distribuição graficamente sob a perspectiva de artista:

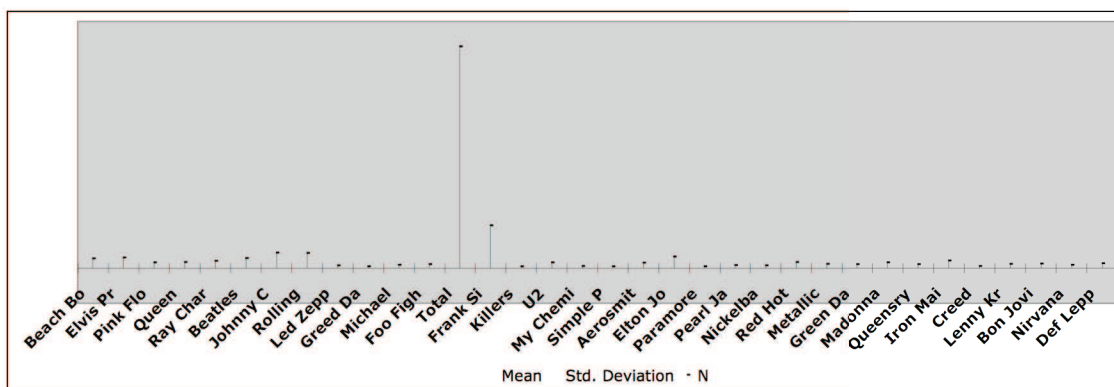


Gráfico 10: Distribuição de artistas na dimensão semântica 1: Ações Pessoais

O gráfico nos mostra os extremos opostos da dimensão, representados pelas bandas Beach Boys (mais representativa) e Def Leppard (menos representativa). Há grande tendência para a diminuição da presença concomitante das características semânticas que denotam movimento, fala, pessoas, tempo e objetos. Ainda assim, é possível destacar a proximidade funcional nas letras das músicas dos Beach Boys, Elvis, Pink Floyd, e assim por diante.

O gráfico seguinte apresenta a distribuição, quando observamos a época de gravação/divulgação das músicas:

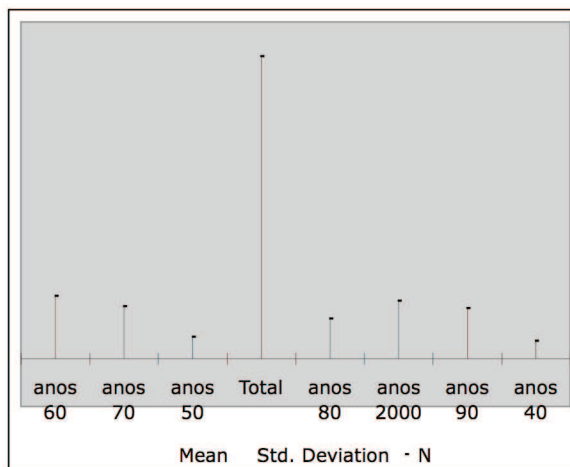


Gráfico 11: Distribuição por período de gravação na dimensão semântica 1: Ações Pessoais

Quando temos o período de gravação/divulgação das letras de música sob perspectiva, percebemos que as letras de músicas dos anos 60, 70 e 50 tendem a enfatizar mais ações pessoais do que as letras de músicas gravadas nos anos 80, 2000, 90 e, especialmente, 40.

A seguir, apresentamos o gráfico que representa a distribuição de letras de música nesta dimensão à luz dos estilos/gêneros musicais.

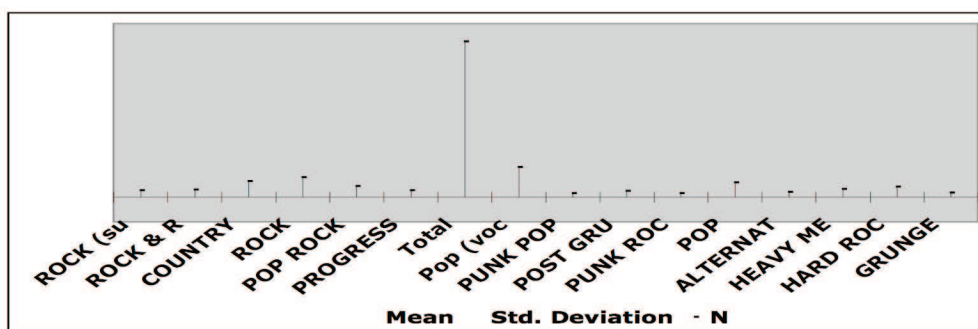


Gráfico 12: Distribuição por período de gravação na dimensão semântica 1: Ações Pessoais

Com a ajuda do gráfico podemos visualizar a distribuição dos estilos musicais nesta dimensão. Em suma, letras de música do estilo “*rock surfista*” mais frequentemente apresentam palavras que nos remetem às ações das pessoas, enquanto letras do estilo “*grunge*” fazem isso menos frequentemente. Nesta dimensão, as letras do estilo *rock’n’roll* estão mais próximas das de *rock surfista* e das de *country*, do que das de

*rock*, por exemplo.

Esta análise responde parcialmente às nossas questões de pesquisa. Esta dimensão apresenta o enfoque semântico que representa uma das características de análise dimensional representadas em relação aos diferentes gêneros musicais, artistas e ao longo do tempo. Veremos, a seguir, a segunda dimensão semântica.

### 5.2.2.2. Segunda Dimensão Semântica: emoções e sociedade

A segunda dimensão semântica é representada por canções que falam sobre emoções e questões sociais, por isso a denominamos “emoções e sociedade”. As associações semânticas que a representam envolvem estados emocionais<sup>167</sup>, processos psicológicos intencionais, capacidades, necessidades, desejos, religião, presenças fantásticas e fantasiosas, drogas, sexo e violência.

Esta dimensão, assim como todas as outras, também foi observada em relação aos artistas, ao tempo e ao estilo das letras das músicas de nosso corpus. A banda Metallica é a mais representativa na dimensão quando consideramos **artistas**, ou seja, a banda apresenta o maior índice de canções que tematizam emoções e questões sociais, como podemos ver no seguinte quadro de estatística descritiva:

Singer_band	Mean	Std.Deviation	N
Metallic	,6407	1,83384	119
Paramore	,4694	1,59044	50
Queen	,3056	1,81314	172
My Chemi	,3049	1,47184	61
Iron Mai	,2707	1,26965	215
Frank Si	,2578	1,72325	1218
Foo Figh	,1867	1,44582	109
Queensry	,1835	1,49291	112
Nirvana	,1260	1,88357	98
Green Da	,0979	1,18411	110
Elvis Pr	,0809	1,61465	305

<sup>167</sup> A lista das palavras incluídas na categoria estados emocionais está apresentada no Anexo 4.

Singer_band	Mean	Std.Deviation	N
Madonna	,0639	1,29933	167
Michael	,0490	1,59906	99
Bon Jovi	,0342	1,37039	127
Nickelba	,0331	1,39783	79
Simple P	,0079	1,53624	49
Total	,0000	1,55867	6290
Lenny Kr	-,0246	1,73601	125
Def Lepp	-,0393	1,34334	142
Aerosmit	-,0437	1,40507	152
Red Hot	-,0540	1,34457	169
Ray Char	-,0757	1,67780	206
U2	-,0853	1,59354	168
Pearl Ja	-,0881	1,62310	85
Elton Jo	-,1173	1,31615	325
Greed Da	-,1428	1,69696	49
Beatles	-,1932	1,80517	281
Killers	-,2467	1,27473	51
Johnny C	-,2942	1,21920	439
Rolling	-,3021	1,53769	430
Creed	-,3066	1,01416	60
Pink Flo	-,3394	1,19370	163
Led Zepp	-,3985	1,12640	79
Beach Bo	-,4227	1,63083	276

Quadro 40: Estatística descritiva para o fator Sem\_C2, em relação a artistas

Metallica e Beach Boys representam os extremos opostos nesta dimensão. Quanto mais próximos das primeiras posições no quadro acima, mais as letras de música desses artistas apresentarão léxico que exprime estados emocionais, processos psicológicos, capacidade, necessidade, religião, presenças fantásticas, drogas, sexo e violência. A música “*Harvester of Sorrow*” é a mais representativa desta categoria. Vejamos a letra:

<p>My life suffocates  Planting seeds of hate  I've loved,  Turned to hate  Trapped far beyond my fate I give  You take  This life that I forsake  Been cheated of my youth  You turned this lie to truth  Anger  Misery  You'll suffer unto me</p>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------



Harvester of sorrow  
 Language of the mad  
 Harvester of sorrow

Pure black looking clear  
 My work is done soon here  
 Try getting back to me  
 Get back what used to be  
 Drink up  
 Shoot in  
 Let the beatings begin  
 Distributor of pain  
 Your loss becomes my gain

Anger  
 Misery  
 You'll suffer unto me

Harvester of sorrow  
 Language of the mad  
 Harvester of sorrow

All have said their prayers  
 Invade their nightmares  
 To see into my eyes  
 You'll find where murder lies Infanticide

Harvester of sorrow  
 Language of the mad  
 Harvester of sorrow  
 Language of the mad  
 Harvester of sorrow

Texto 39: Letra da música *Harvester of sorrow*– gravação do Metallica, 1988

Essa música está no topo da categoria por apresentar-se rica em expressões que exprimem sentimento, tais como, “*love*”, “*hate*” e “*sorrow*”, ou seja, apresenta emoções de amor, raiva, sofrimento e tristeza; focaliza a questão da violência (“*beatings*”, “*loss*”) e traz referências psicológicas (“*mad*”; em português, louco). Assim, representa com clareza esta dimensão.

No que diz respeito ao **tempo**, as músicas gravadas nos anos 40 são as mais representativas, enquanto as dos anos 70 são as menos representativas, como demonstra o quadro a seguir:

### Descriptive Statistics

Dependent Variable: Sem\_C\_2

Year	Mean	Std. Deviation	N
4	,3108	1,70003	367
5	,2534	1,84882	453
8	,0835	1,53438	832
0	,0756	1,41560	1200
9	,0597	1,51546	1046
Total	,0000	1,55867	6290
6	-,1682	1,66029	1305
7	-,2136	1,40868	1087

Quadro 41: Estatística descritiva para o fator Sem\_C2, em relação ao período de gravação

A canção “I believe”, gravada por Frank Sinatra em 1943 é a mais representativa.

Vejamos:

I believe for every drop of rain that falls, a flower grows.  
 I believe that somewhere in the darkest night, a candle glows.  
 I believe for everyone who goes astray, someone will come to show the way  
 I believe, I believe.  
 I believe above the storm the smallest prayer will still be heard.  
 I believe that someone in the great somewhere hears every word.  
 Every time I hear a newborn baby cry, or touch a leaf or see the sky,  
 Then I know why I believe!

Texto 40: Letra da música *I believe*– gravação de Frank Sinatra, 1943

A música é marcada pela palavra do título, “acredito” (“*believe*”), que compreende processos psicológicos, necessidade e religião. Religião também está marcada por oração e vela (“*prayer*” e *candle*”). As emoções representadas na letra da canção são mais positivas em relação à anterior, mas a dimensão não leva em conta a positividade ou não de emoções, apenas a presença de léxico relacionado ao emocional.

O **estilo** que mais salienta questões emocionais e sociais é o *Heavy Metal*, em oposição ao estilo *rock surfista*, como podemos ver a seguir:

### Descriptive Statistics

Dependent Variable: Sem\_C\_2

Style	Mean	Std. Deviation	N
HEAVY ME	,4025	1,50304	334
PUNK POP	,2654	1,53088	160
Pop (voc	,2578	1,72325	1218
ROCK & R	,0809	1,61465	305
GRUNGE	,0266	1,76586	183
PUNK ROC	,0237	1,36182	159
POST GRU	,0184	1,34720	248
Total	,0000	1,55867	6290
HARD ROC	-,0187	1,37121	421
POP	-,0383	1,36333	591
ALTERNAT	-,0987	1,32834	220
PROGRESS	-,1264	1,34597	275
ROCK	-,1389	1,61575	806
POP ROCK	-,1528	1,72800	449
COUNTRY	-,2244	1,38472	645
ROCK (su	-,4227	1,63083	276

Quadro 42: Estatística descritiva para o fator Sem\_C2, em relação ao gênero/estilo musical

A classificação estilística é correspondente à classificação por artista, ou seja, a banda Metallica, artista mais representante desta dimensão, pertence ao estilo *heavy metal* e a banda Beach Boys ao estilo rock surfista. Todavia, quando ao estilo *heavy metal* são incorporadas as letras de todos os artistas representantes do estilo em nosso corpus, existem letras de música que, individualmente, apresentam escores superiores às do Metallica, isto é, as letras mais representativas mudam um pouco de posição no que diz respeito a seus escores médios. Seleccionamos a letra da canção “*Moonchild*”, da banda de *heavy metal* inglesa Iron Maiden, gravada em 1988, para observar nesta análise:

```

Seven deadly sins
Seven ways to win
Seven holy paths to hell
And your trip begins

Seven downward slopes
Seven bloodied hopes
Seven are your burning fires,
Seven your desires

I am he the bornless one
The fallen angel watching you
Babylon, the scarlet whore
I'll infiltrate your gratitude
Don't you dare to save your son
Kill him now and save the young ones
Be the mother of a birth strangled babe
Be the devils own, Lucifer's my name

```

Moonchild - hear the mandrake scream  
 Open the seventh seal  
 Moonchild - You'll be mine soon child  
 Moonchild - take my hand tonight  
  
 I count the heads of those unborn  
 The accursed ones I'll find them all  
 If you die by your own hand  
 As a suicide you shall be damned  
 And if you try to save your soul  
 I will torment you - you shall not grow old  
 With every second and passing breath  
 You'll be so alone your soul will bleed to death  
  
 The twins they are exhausted, seven is the night  
 Gemini is rising as the red lips kiss to bite  
 Seven angels seven demons battle for his soul  
 When Gabriel lies sleeping, this child was born to die  
  
 One more dies one more lives  
 One baby cries One mother grieves  
 For all the sins you will commit  
 You'll beg forgiveness and none I'll give  
 A web of fear shall be your coat  
 To clothe you in the night  
 A lucky escape for you youngman  
 But I see you damned in endless night.

Texto 41: Letra da música *Moonchild*– gravação de Iron Maiden, 1988

A canção enfoca questões emocionais em relação ao fantástico religioso, principalmente, pois fala dos pecados capitais, de céu e inferno, de esperança e perdão, de demônios e anjos, de sofrimento e morte. Retrata também problemas enfrentados pelas pessoas como a morte, o suicídio, a prostituta e o medo.

Como vimos, uma segunda dimensão de análise semântica categoriza letras de música pela presença de palavras que denotam emoções tipicamente, mas não exclusivamente, negativas normalmente associadas a outras questões presentes em nossas interações sociais, como religião, fenômenos (e “presenças”) fantásticas e fantasiosas, problemas sociais que envolvem violência e drogas; além de tratarem dos desejos, capacidades e pensamentos dos envolvidos.

O quadro abaixo mostra como a presença de léxico com essas referências se distribui mais ou menos marcadamente em relação aos artistas presentes em nosso corpus:

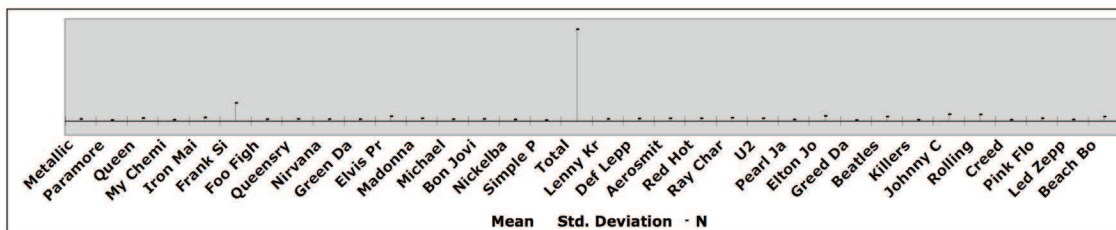


Gráfico 13: Distribuição de artistas na dimensão semântica 2: Emoções e Sociedade

O gráfico demonstra que, quando observamos o léxico que se refere a emoções e questões sociais, letras de música de artistas como Metallica, Paramore, Queen, My Chemical Romance, Iron Maiden e Frank Sinatra, por exemplo, assemelham-se, apresentando maior ocorrência dessa temática que as letras de música de bandas como Beach Boys, Led Zeppelin e Pink Floyd. Em relação ao estilo/gênero musical a distribuição lexical nesta categoria pode ser observada no gráfico apresentado a seguir.

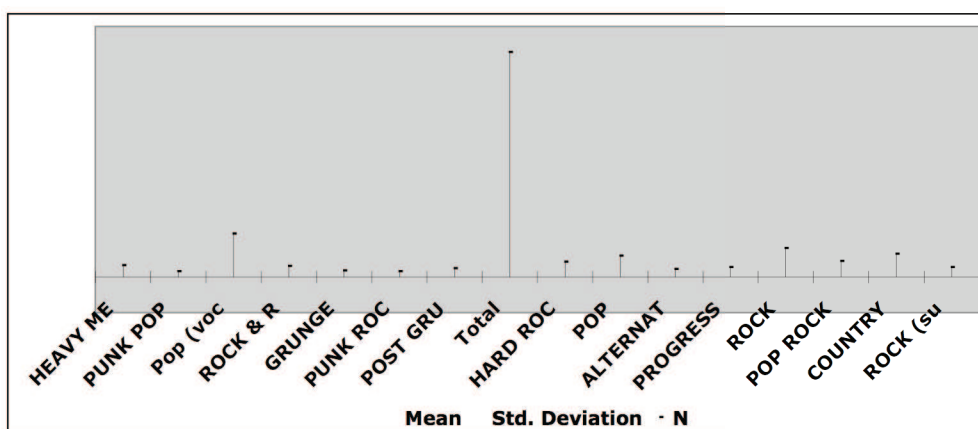


Gráfico 14: Distribuição por estilo/gênero musical na dimensão semântica 2: Emoções e Sociedade

Como vemos, quando agrupadas todas as letras de música gravadas por cada um dos artistas categorizados sob cada um dos quinze estilos estudados, o *heavy metal* demonstra maior preocupação emocional social. Essa característica vai se diluindo, isto é, diminuindo, conforme vamos observando o gráfico para a direita. Em outras palavras, o estilo *punk pop* apresenta léxico referente à emoção e sociedade em menor proporção que o estilo *heavy metal*, mas em maior proporção que o estilo *pop vocal*, que, por sua vez, apresenta maior proporção que o estilo *rock'n'roll*, que apresenta maior proporção que o estilo *grunge*, e assim por diante, até chegar ao estilo *rock surfista*, que é o que

apresenta menor ocorrência de itens lexicais que se referem à emoção e questões sociais ao mesmo tempo.

Em relação ao período de gravação das músicas, essa distribuição acontece da seguinte maneira:

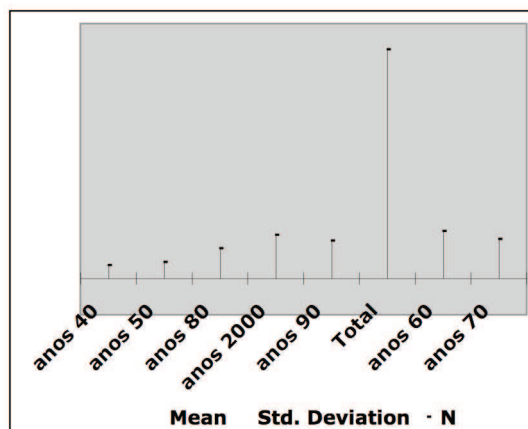


Gráfico 15: Distribuição por período de gravação na dimensão semântica 2: Emoções e Sociedade

Músicas gravadas nos anos 60 e 70 apresentam menos referências lexicais a estados emocionais relacionados a questões sociais, do que as letras de música gravadas nos outros períodos abordados em nosso corpus de estudo, segundo o gráfico anterior.

A posição no extremo oposto ao mais marcante na dimensão nos gráficos apresentados não significa que naquele determinado período (anos 60 e 70), ou aqueles gêneros musicais (*hard rock, pop, alternative rock, progressivo, rock, pop rock, country e rock surfista*) ou ainda aqueles artistas (Lenny Kravitz, Def Leppard, Aerosmith, Red Hot Chilly Peppers, Ray Charles, U2, Pearl Jam, Elton John, Greenday, Beatles, Killers, Johnny Cash, Rolling Stones, Creed, Pink Floyd e Beach Boys) não tenham representações lexicais de estados emocionais, processos psicológicos, desejo, capacidade, necessidade, religião, violência, drogas, sexo ou presenças fantásticas e fantasiosas. Outrossim, significa que esses referenciais ocorrem simultaneamente menos frequentemente. Isto é, esses artistas, períodos e estilos podem ter letras de música que falam exclusivamente de emoções voltadas para o amor, ou de emoções voltadas para o ódio, ou letras de música que falam sobre a violência, ou sexo; por outro lado suas letras de música pouco vão agrupar todas essas características.

### 5.2.2.3. Terceira Dimensão Semântica: referência a música

A terceira dimensão semântica classifica as letras de música que falam sobre música, ou seja, que contém palavras que se referem a música, instrumentos e estilos musicais, dança, diferentes ritmos e sons.

A dimensão é fortemente representada por Frank Sinatra, tanto enquanto artista, como por período de tempo (anos de 40), quanto por estilo musical (*pop vocal*). O quadro seguinte apresenta o escore médio por **artista**, obtido pela extração fatorial.

Singer_band	Mean	Std. Deviation	N
Frank Si	,2568	1,48683	1218
Beach Bo	,1807	1,53799	276
Madonna	,1385	,76718	167
Ray Char	,1096	,89190	206
Red Hot	,0484	,60649	169
Michael	,0188	,99078	99
U2	,0027	,91632	168
Total	,0009	1,00373	6290
Rolling	,0006	1,18967	430
Johnny C	,0000	,84952	439
Beatles	-,0292	1,05588	281
Lenny Kr	-,0769	1,51671	125
Aerosmit	-,0770	,60595	152
Bon Jovi	-,0798	,42636	127
Elton Jo	-,0839	,56710	325
Paramore	-,0896	,76873	50
Pink Flo	-,0918	,51232	163
Queen	-,0990	,67408	172
Creed	-,1054	,45995	60
Led Zepp	-,1101	,41061	79
Nirvana	-,1198	,60098	98
Elvis Pr	-,1223	,73561	305
Killers	-,1341	,39981	51
My Chemi	-,1756	,35773	61
Iron Mai	-,1788	,53782	215
Green Da	-,1885	,44888	110
Metallic	-,2013	,40322	119
Foo Figh	-,2051	,58050	109
Def Lepp	-,2107	,29784	142

Greed Da	-,2228	,28638	49
Queensry	-,2257	,27150	112
Nickelba	-,2319	,33268	79
Pearl Ja	-,2784	,23579	85
Simple P	-,3093	,15544	49

Quadro 43: Estatística descritiva para o fator Sem\_C3, em relação a artistas

O quadro demonstra que as letras de músicas gravadas por Frank Sinatra, Beach Boys e Madonna, por exemplo, apresentam mais palavras que se referem à música do que as letras de música de artistas como Simple Plan, Pearl Jam ou Nickleback. O artista que apresenta mais referenciais musicais em suas letras é Frank Sinatra. Apresentaremos, a seguir, a análise de duas gravações de Frank Sinatra que são as mais representativas nesta dimensão. A primeira, “*I write the songs*”, foi gravada em 1994 e é a mais representativa tanto quando observamos o artista, como quando observamos o estilo mais representativo, o *pop vocal*.

I've been alive forever,  
 And I sang the very first song,  
 I put the words and the melodies together,  
 I am music and I sing this song:  
 I sing the songs that make the whole world sing,  
 I sing the songs of love and special things,  
 I sing the songs that make the young girls cry,  
 I sing the songs, I sing the songs.  
 My home is deep within you,  
 And I've got my own place in your soul,  
 Now when I look out through your eyes,  
 I am young again even though I'm growing old.  
 I sing the songs that make the whole world sing,  
 I sing the songs of love and special things,  
 I sing the songs that make the young girls cry,  
 I sing the songs, I sing the songs.  
 My music makes you dance,  
 It gets your spirit to take a chance,  
 And if the music's from the heart,  
 It's a real fine way to start,  
 It's from me, it's to you,  
 It's from you, it's to me,  
 It's a world-wide symphony.  
 I sing the songs that make the whole world sing,  
 I sing the songs of love and special things,  
 I sing the songs that make the young girls cry,  
 I sing the songs, I sing the songs.  
 I sing, and the whole world sings,  
 I sing the songs of love and special things,  
 I sing the songs that make the young girls cry,  
 I sing the songs, I sing the songs.  
 I am music, and I sing the songs.

Texto 42: Letra da música *I write the songs*– gravação de Frank Sinatra, 1994



Nitidamente a música é representante desta categoria pela repetição de “*I sing the song*”, mas também apresenta outras palavras que representam a dimensão como “*music*”, “*dance*” e o passado de cantar “*sang*”. Em suma, é uma música cuja letra reflete a importância da música, do ato de cantar, para o autor. A seguir, apresentamos o quadro com a estatística descritiva em relação ao **período** de gravação das letras de música em nosso corpus, para esta dimensão.

Year	Mean	Std. Deviation	N
4	,2043	1,25921	367
5	,0960	1,13383	453
6	,0937	1,30274	1305
7	,0650	1,11004	1087
Total	,0009	1,00373	6290
8	-,0321	,82515	832
0	-,1120	,56894	1200
9	-,1385	,75648	1046

Quadro 44: Estatística descritiva para o fator Sem\_C3, em relação ao período de gravação

Quando as letras de música são agrupadas por período de gravação, a década que mais apresenta elementos lexicais referentes à manifestação musical é a década de 40. Em nosso corpus, esse período é composto exclusivamente por gravações de Frank Sinatra. A letra da música “*If loveliness were music*”, gravada em 1946, representa esta categoria de análise. Vejamos:

```

If loveliness were music
I'd hear a symphony every time I look at you
If loveliness were music
I know your charms would be heavenly to listen to
I'd hold you near, and then I'd hear the ringing, singing
Of themes from angels' dreams

If loveliness were music
The sight of you would fill
All the world with songs that thrill
Earth and heaven too
Would echo songs of you
I'd hold you near, and then I'd hear the ringing, singing
Of themes from angels' dreams

If loveliness were music
The sight of you would fill
All the world with songs that thrill
Earth and heaven too
Would echo songs of you

```

Texto 43: Letra da música *If loveliness were music*– gravação de Frank Sinatra, 1946

Essa letra de música apresenta léxico referente a elementos musicais como

“*symphony*”, “*hear*”, “*listen*”, “*ringing*”, “*singing*”, “*music*” e “*songs*”. O autor compara a presença da amada ao prazer de ouvir uma sinfonia.

As tabelas de análise descritiva para esta dimensão apontam que as músicas menos representativas entre os artistas são as gravadas pela banda canadense Simple Plan, sob a perspectiva de tempo, as músicas menos representativas são as gravadas nos anos 90 e entre os estilos, são as de *punk rock*. O quadro seguinte mostra a estatística descritiva em relação ao **estilo** para esta dimensão.

Style	Mean	Std. Deviation	N
Pop (voc	,2568	1,48683	1218
ROCK (su	,1807	1,53799	276
COUNTRY	,0350	,86410	645
ALTERNAT	,0061	,56976	220
Total	,0009	1,00373	6290
POP	-,0038	,71736	591
POP ROCK	-,0173	1,00500	449
ROCK	-,0435	1,10624	806
ROCK & R	-,1223	,73561	305
HARD ROC	-,1229	,46931	421
PROGRESS	-,1463	,43518	275
HEAVY ME	-,1868	,49357	334
POST GRU	-,1895	,48501	248
PUNK POP	-,1896	,49531	160
GRUNGE	-,1935	,47376	183
PUNK ROC	-,1990	,40518	159

Quadro 45: Estatística descritiva para o fator Sem\_C3, em relação ao estilo/gênero musical

O quadro aponta que o estilo musical que mais apresenta letras de música que falam sobre música é o *pop vocal*, como dissemos anteriormente. Da mesma forma, o *punk rock* é o estilo/gênero musical que tem menos letras de música que falam sobre música. Apresentamos, a seguir, uma amostra de canção não representativa desta dimensão, a qual é a menos representativa por estilo, a gravação de 2009, pela banda Greenday (*punk rock*), da canção intitulada “*Restless Heart Syndrome*”:

<p>I've got a really bad disease  It's getting begging on my hands and knees  Take me to the emergency  Cause something seems to be missing  Somebody take the pain away  It's like an ulcer bleeding in my brain  Send me to the pharmacy so I can lose my memory</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

I'm elated, medicated  
 Lord knows I tried to find a way to run away

I think they found another cure  
 For broken hearts and feeling insecure  
 You'd be surprised what I endure  
 What makes you feel so self-assured

I need to find a place to hide  
 You never know what could be waiting outside  
 The accidents that you could find  
 It's like some kind of suicide

So what kills you, is what impales you  
 I feel like I've been crucified to be satisfied

I'm elated, medicated  
 I am my own worst enemy  
 So what ails you is what impales you  
 You are your own worst enemy  
 You're a victim of the system

I'm a victim of my symptom  
 I am my own worst enemy  
 You're a victim of your symptom  
 You are your own worst enemy  
 Know your enemy

Texto 44: Letra da música *Restless Heart Syndrome*– gravação do Greenday, 2009

Assim como diversas outras letras de música, essa não apresenta nenhuma associação a manifestação musical sendo, portanto, não representativa desta dimensão. Entre as letras de música mais e menos representativas desta dimensão ficam as outras que, conforme contenham mais itens lexicais referentes à música, mais se aproximam da primeira. A distribuição de representatividade em relação a artistas se dá da seguinte maneira:

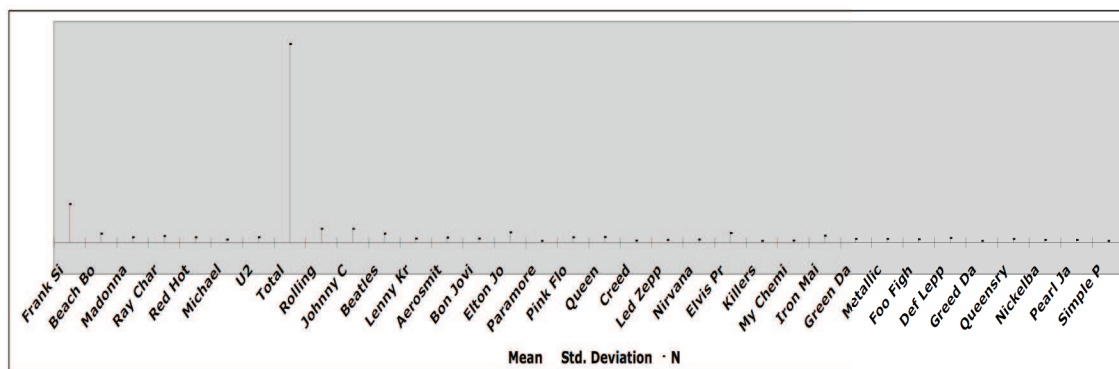


Gráfico 16: Distribuição de artistas na dimensão semântica 3: referência a música

O gráfico demonstra que os artistas que mais falam sobre música, quando tem a

totalidade de suas gravações observadas são Frank Sinatra, Beach Boys, Madonna, Ray Charles, Michael Jackson e U2. Já artistas como Simple Plan, Pearl Jam, Nickelback, Queensryche Greenday, Def Leppard, entre outros, apresentam menos referências musicais em suas letras. O quadro também demonstra que, quando as letras são consideradas por artistas, a presença de elementos que representam a manifestação musical é baixa, pois a maioria dos artistas analisados encontra-se abaixo do escore médio. Quando observadas de acordo com o estilo musical, as letras de música se distribuem da seguinte maneira:

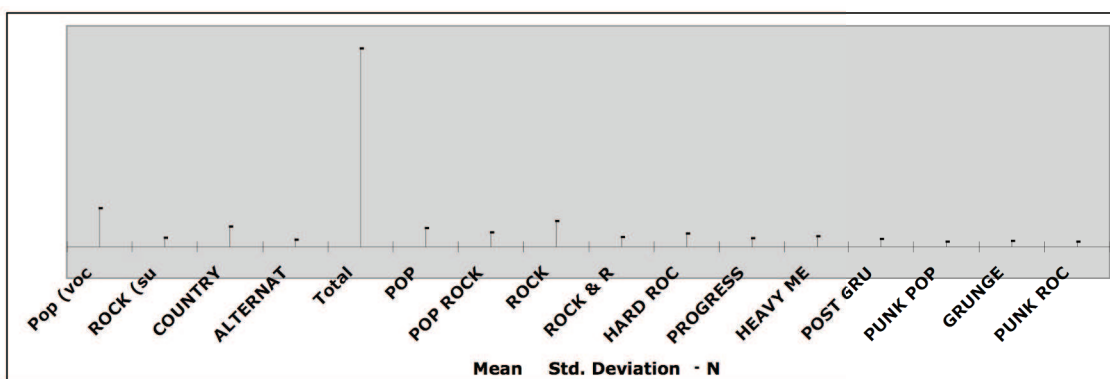


Gráfico 17: Distribuição de período de gravação na dimensão semântica 3: referência a música

O gráfico demonstra que quando agrupadas por estilo, as letras de música mais representativas de manifestações musicais são as dos estilos *pop vocal*, *rock surfista*, *country* e *alternativo*, enquanto as menos representativas são a do estilo *punk rock*, *grunge* e *punk pop*. Ainda que apresentem baixa ocorrência lexical que se refira à música, não significa que esses artistas nunca falem sobre música em suas letras.

Por outro lado, quando a manifestação musical é observada em relação ao tempo, a distribuição é menos desequilibrada, como mostra o gráfico a seguir:

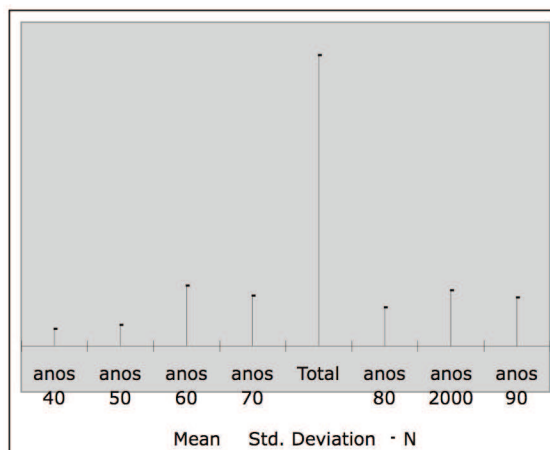


Gráfico 18: Distribuição de período de gravação na dimensão semântica 3: referência a música

O gráfico de distribuição por período de gravação mostra que nas últimas décadas há menos referências musicais nas letras de música do que antes nos anos 40 e 70, quando observamos os gêneros/estilos musicais.

O contraste de dados entre os últimos gráficos denota que ainda que um artista possa ser categorizado entre os menos representativos de uma determinada dimensão, as letras das músicas por ele gravadas e que apresentam os elementos da dimensão serão consideradas, quando observadas por período ou estilo musical. O cantor Elvis Presley, por exemplo, está caracterizado como pouco representativo desta dimensão de manifestação musical, mas suas gravações vão de 1956 a 1977, dois períodos que são considerados entre os mais representativos aqui. Um exemplo disso é a letra da música “*Rip it up*”, gravada pelo cantor em 1956.

```
'Cause it's Saturday nite and I just got paid
Fool about my money don't try to save
My heart says go, go
Have a time 'cause it's Saturday nite
And I'm feelin' fine

I'm gonna rip it up
I'm gonna rock it up
I'm gonna shake it up
I'm gonna ball it up
I'm gonna rip it up
And ball tonite

I got a date and I won't be late
Pick her up in my '88'
Shag it on down to the union hall
```

When the music starts jumpin'  
I'll have a ball

Along about 10 I'll be flying high  
Rocking on out into the sky  
I don't care if I spend my gold  
'Cause tonite I'm gonna be one happy soul

Texto 45: Letra da música *Rip it up*– gravação de Elvis Presley, 1956

A letra de música apresenta elementos musicais referentes à dança, tais como “rock”, “shake”, “ball” e “rocking”, além da palavra “music”. Quando analisada individualmente pela presença de léxico que se refere a música, essa letra apresenta um escore de 6,57, longe de ser zero. Vale ressaltar que o escore da música “*I write the song*”, a mais representativa desta dimensão é 13,36.

Vimos que, em termos semânticos podemos observar letras de música sob três dimensões principais, músicas que falam sobre as ações das pessoas, músicas que falam sobre emoções e problemas sociais, e músicas que falam sobre música. O quadro abaixo apresenta um resumo das dimensões semânticas.

DIMENSÃO		Ações Pessoais	emoções e sociedade	referência a musica
relações semânticas de		movimento, fala, pessoas, tempo e objetos	estados emocionais, processos psicológicos e intencionais, capacidade, necessidades, desejos, religião, presenças fantásticas e fantasiosas, drogas, sexo e violência.	música, instrumentos e estilos musicais, dança, diferentes ritmos e sons.
mais representativos	artista	Beach Boys	Metallica	Frank Sinatra
	período	1960	1940	1940
	estilo	Rock Surfista	Heavy Metal	Pop Vocal
escore médio	artista	Foo Fighters	Simple Plan	Eolling Stones
	período	1980	1990	1980
	estilo	Punk Pop	Post Grunge	Pop
menos representativos	artista	Def Leppard	Beach Boys	Simple Plan
	período	1940	1960	1990
	estilo	Grunge	Rock Surfista	Punk Rock

Quadro 46: Resumo das dimensões semânticas

A análise fatorial descrita no capítulo metodológico resultou em três fatores para três esferas diferentes: gramatical, semântica e léxico-gramatical. A interpretação desses fatores apontou para três dimensões de análise para cada uma das esferas de

características, que foram observadas sob três diferentes perspectivas, isto é, em relação aos artistas que gravaram as músicas, ao período em que foram gravadas e ao estilo musical em que esses artistas foram categorizados. Até o momento, apresentamos as dimensões gramaticais e semânticas. Na seção seguinte, apresentamos as dimensões léxico-gramaticais.

### **5.2.3. Dimensões Léxico-Gramaticais**

Em nossa análise interpretativa dos fatores léxico-gramaticais (seção 5.1.7) demonstramos que existe pouquíssima variação entre esses fatores e os puramente gramaticais. Essa baixa diferença reside no fato de serem consideradas para a esfera léxico-gramatical as mesmas variáveis consideradas para a esfera gramatical, acrescidas das contagens de frequência dos trigramas mais típicos de letras de música no corpus e no inglês geral online, como também já foi dito. Consequentemente, existe pouca diferença entre as dimensões gramaticais e as léxico-gramaticais. Tão pouca que conferimos a mesma nomenclatura às segunda e terceira dimensões. Isso não significa que essas dimensões não devam ser consideradas. Pelo contrário, elas existem para oferecer ao pesquisador uma opção quanto às características que pretende investigar em suas letras. Se sua investigação objetivar uma análise puramente gramatical, então o pesquisador pode considerar apenas as dimensões gramaticais. Todavia, se o pesquisador quiser observar o comportamento gramatical associado à padronização linguística do inglês geral, ele deve considerar as dimensões léxico-gramaticais.

Levando essa semelhança em consideração, não apresentaremos uma análise detalhada para cada dimensão léxico-gramatical, pois seria uma repetição das análises que apresentamos para as dimensões gramaticais, feitas no item 5.2 deste trabalho. Primeiramente, discutiremos de maneira geral as três dimensões léxico-gramaticais. Em seguida, apresentaremos alguns pontos diferenciais entre a primeira dimensão léxico-gramatical e a primeira dimensão gramatical. Finalmente, apresentaremos um quadro de resumo para as dimensões semânticas, que explicitará nossa discussão.

As três dimensões léxico-gramaticais por nós identificadas foram:

Dimensão 1 – Argumentação e Padronização

Dimensão 2 – Interação e Descrição

Dimensão 3 – Narrativa no passado X referencial imediato

A primeira dimensão é caracterizada pelos elementos gramaticais de cunho argumentativos, representados por infinitivos, gerúndios, imperativos e modais, associados aos trigramas mais frequentes, em oposição à presença de substantivos.

A segunda dimensão léxico-gramatical é caracterizada pela presença simultânea de pronomes pessoais e possessivos em oposição à presença simultânea de adjetivos e determinantes, que caracterizam a oposição entre a função interativa e a função descritiva do discurso.

A terceira dimensão léxico-gramatical é marcada pela oposição entre verbos no passado (que caracterizam um discurso mais narrativo no passado) e verbos no presente (que denotam o imediatismo do discurso).

A seguir, apresentamos de modo sucinto as análises descritivas apenas para a primeira dimensão gramatical, considerando sua característica peculiar quanto à presença de trigramas. As tabelas de estatística descritiva extraídas para essa dimensão revelam as seguintes características:

a) por **artista**: Simple Plan (+) X Aerosmith (-)

**Descriptive Statistics**

Dependent Variable: LexGram\_C\_1

Singer_band	Mean	Std. Deviation	N
Simple P	2,9234	3,10550	49
Paramore	2,3146	3,53101	50
Lenny Kr	1,9424	3,71916	125
Beatles	1,5725	3,47611	281
Creed	1,4401	2,53287	60
Madonna	1,4172	2,84082	167
Nickelba	1,2463	2,98947	79
My Chemi	1,2033	2,38257	61
Michael	,6696	2,67374	99
Foo Figh	,5637	3,00148	109
Elvis Pr	,4737	2,95439	305



Killers	,4721	2,50592	51
Bon Jovi	,4643	2,35203	127
Def Lepp	,4084	3,00155	142
Led Zepp	,3451	2,49396	79
Greed Da	,3353	3,05296	49
Queen	,2301	2,89884	172
Nirvana	,1112	3,56448	98
Total	,0263	3,05692	6290
Ray Char	-,0030	2,90781	206
U2	-,0059	2,98633	168
Rolling	-,0334	3,09902	430
Frank Si	-,0342	3,09086	1218
Red Hot	-,3051	2,88514	169
Queensry	-,4430	2,13716	112
Beach Bo	-,5201	3,45310	276
Green Da	-,6310	2,85103	110
Pink Flo	-,7806	3,02239	163
Iron Mai	-,8060	2,82737	215
Elton Jo	-,8936	2,41367	325
Johnny C	-,9444	2,47903	439
Pearl Ja	-1,0396	3,07726	85
Metallic	-1,1017	2,44516	119
Aerosmit	-1,2073	2,64873	152

Quadro 47: Estatística descritiva para o fator LexGram\_C1,  
em relação ao artista

b) por **período**: anos 2000 (+) X anos 70 (-)

**Descriptive Statistics**  
Dependent Variable: LexGram\_C\_1

Year	Mean	Std. Deviation	N
0	,8285	2,95535	1200
5	,1627	3,18331	453
6	,1322	3,25973	1305
4	,0706	3,05372	367
Total	,0263	3,05692	6290
9	-,2337	3,03751	1046
8	-,4383	2,87860	832
7	-,4522	2,81811	1087

Quadro 48: Estatística descritiva para o fator LexGram\_C1,  
em relação ao período de gravação

c) por **estilo**: *Punk Pop* (+) X *Heavy Metal* (-)

**Descriptive Statistics**  
Dependent Variable: LexGram\_C 1

Style	Mean	Std. Deviation	N
PUNK POP	2,0774	3,07039	160
POST GRU	,9932	2,90588	248
POP ROCK	,9819	3,38543	449
ROCK & R	,4737	2,95439	305
ROCK	,3663	3,17987	806
Total	,0263	3,05692	6290
POP	,0212	2,78216	591
Pop (voc	-,0342	3,09086	1218
ALTERNAT	-,1250	2,81555	220
HARD ROC	-,1581	2,79832	421
PUNK ROC	-,3332	2,93930	159
GRUNGE	-,4233	3,38724	183
ROCK (su	-,5201	3,45310	276
PROGRESS	-,6431	2,69793	275
COUNTRY	-,6438	2,65787	645
HEAVY ME	-,9114	2,69741	334

Quadro 49: Estatística descritiva para o fator LexGram\_C1, em relação ao estilo/gênero

Esses dados, quando comparados aos dados obtidos pela estatística descritiva na dimensão gramatical 1, “Argumentação versus Informação”, revelam que não há diferença em se tratando de artistas, pois naquele os polos opostos também estão marcados pelas bandas Simple Plan (+) e Aerosmith (-). Porém, no que diz respeito a décadas e estilos, vemos uma leve diferença.

Enquanto a análise puramente gramatical revela oposição entre as décadas de 2000 (+) e de 80 (-), a dimensão léxico-gramatical revela oposição entre os períodos de 2000 (+) e 1970 (-). Essas características são representadas principalmente pela música “*My number one*”, de 2009, da banda Paramore, que apresentamos no item 5.1.7. A letra demonstra que autora tenta persuadir o interlocutor a conhecer suas qualidades, usando expressões comuns à música e ao inglês geral, como “*oh oh oh*” e “*I love you*”, além de usar outros elementos argumentativos, como já vimos.

O polo oposto é representado pela música “*Shortenin’ Bread*”, gravada em 1979, pelo conjunto Beach Boys:

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread  
 Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

Put on the skillet  
 Slip on the lid  
 Mammy's gonna make us some shortenin' bread  
 And that ain't all  
 Our mammy's gonna do  
 She's gonna cook us some coffee, too

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread  
 Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

I slipped in the kitchen  
 Raised up the lid  
 I stole me a mess o' that shortenin' bread  
 I walked up to a pretty girl and I said  
 Baby how'd you like some shortenin' bread

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread  
 Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

They caught me with the skillet  
 They caught me with the lid  
 They caught me with the girl eatin' shortenin'  
 bread  
 Six months for the skillet  
 Six months for the lid  
 Now I'm doin' time for eatin' shortenin' bread

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread  
 Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

Shortenin'  
 Shortenin' bread  
 Shortenin'  
 Shortenin' bread

Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread  
 Mama's little baby loves shortenin' shortenin'  
 Mama's little baby loves shortenin' bread

Texto 46: Letra da música *Shortenin' Bread*– gravação dos Beach Boys, 1979

A música não apresenta características argumentativas, mas descritivas, pois é rica em substantivos. Além disso, a música apresenta construções não típicas de música nem

do inglês geral, isto é, dos 242 trigramas que essa letra de música possui, apenas 10 aparecem no corpus do Google N-grams mais de 50.000 vezes (“*a pretty girl*”; “*and I said*”; “*in the kitchen*”; “*months for the*”; “*put on the*”; “*six months for*”; “*up to a*”; “*walked up to*”; “*with the girl*”; “*you like some*”), o que indica para um distanciamento da língua inglesa padronizada.

Já a esfera de estilos musicais para a dimensão puramente gramatical aponta oposição entre *Punk Pop* e *Rock Surfista*, mas, quando acrescentadas as contagens de trigramas, ou seja, para a dimensão léxico-gramatical, essa oposição se transforma em *Punk Pop (+)* versus *Heavy Metal (+)*.

A música mais representativa do estilo *Punk Pop* é também “*my number one*”, da banda Paramore, já analisada. No polo negativo a música mais representativa é “*Memory Remains*”, da banda Metallica.

```

Fortune, fame
Mirror vain
Gone insane
But the memory remains

Heavy rings on fingers wave
Another star denies the grave
See the nowhere crowd
Cry the nowhere cheers of honor

Like twisted vines that grow
Hide and swallow mansions whole
And dim the light of an already
Faded prima donna

Fortune, fame
Mirror vain
Gone insane...
Fortune, fame
Mirror vain
Gone insane...
But the memory remains

Heavy rings hold cigarettes
Up to lips that time forgets
While the Hollywood sun sets
Behind your back

And can't the band play on
Just listen, they play my song
Ash to ash
Dust to dust
Fade to black

```

```

Fortune, fame
Mirror vain
Gone insane...
Fortune, fame
Mirror vain
Gone insane...
Dance little tin goddess

Drift away
Fade away
Little tin goddess
Ash to ash
Dust to dust
Fade to black

Fortune, fame
Mirror vain
Gone insane...
Fortune, fame
Mirror vain
Gone insane...
But the memory remains

Ash to ash
Dust to dust
Fade to black...
But the memory remains

Dance little tin goddess dance

```

Texto 47: Letra da música *Memory Remains*– gravação dos Metallica, 1997

A música é fortemente marcada pelo uso de substantivos, sem grandes marcações argumentativas, ou seja, poucos infinitivos (“*see*”, “*cry*”, “*dance*”), apenas um modal (“*can’t*”) e nenhum gerúndio. Dos 171 trigramas que foram gerados para essa letra de música, apenas 5 são representativos do inglês geral, isto é, aparecem mais de 50.000 vezes no corpus do Google N-Grams (“*behind your back*”; “*fade to black*”; “*of an already*”; “*the light of*”; “*you back and*”).

Como pudemos observar, embora bastante semelhante à primeira dimensão gramatical, a primeira dimensão léxico gramatical “argumentação e padronização” também é caracterizada pela presença, ou ausência, de elementos padronizados da língua inglesa em geral.

Vimos que a análise léxico-gramatical é muito semelhante à análise puramente gramatical, mas nem por isso deixa de ser importante. A escolha de quais características devem ser privilegiadas, embora seja subjetiva, cabe ao pesquisador.

O quadro seguinte apresenta um resumo das dimensões léxico-gramaticais:

<b>DIMENSÃO</b>		<b>ARGUMENTAÇÃO E PADRONIZAÇÃO</b>	<b>INTERAÇÃO X DESCRIÇÃO</b>	<b>NARRATIVA PASSADA X REFERENCIAL IMEDIATO</b>
características		infinitivos; gerúndios; imperativos; modais e trigramas mais frequentes nas músicas e no inglês geral X substantivos	pronomes pessoais e possessivos X adjetivos e determinantes	verbos no passado X verbos no presente
polo positivo	artista	Simple Plan	Madonna	Johnny Cash
	período	anos 2000	anos 40	anos 70
	estilo	Punk Pop	Rock'n'roll	country
polo negativo	artista	Aerosmith	Pink Floyd	Metallica
	período	anos 70	anos 70	anos 40
	estilo	Heavy Metal	progressivo	Heavy Metal

Quadro 50: Resumo das dimensões léxico-gramaticais

#### 5.2.4. Síntese dos resultados

Neste capítulo, apresentamos as dimensões de variação linguística na música popular anglo-americana. A distribuição dimensional observou as músicas em termos de artista, década de gravação e estilo musical. As dimensões partiram de perspectivas gramaticais, semânticas e léxico-gramaticais, as quais apontaram, cada uma, para três dimensões de análise.

Quando consideramos características de cunho gramatical as letras de música podem ser analisadas sob as seguintes dimensões:

Dimensão Gramatical 1: Argumentação X Informação

Dimensão Gramatical 2: Interação X Descrição

Dimensão Gramatical 3: Narrativa no Passado X Referencial Imediato

Quando observadas por suas características semânticas, ou seja, pelos assuntos mais representados pela frequência lexical, as letras de música podem ser analisadas sob as seguintes dimensões:

Dimensão Semântica 1: Ações Pessoais

Dimensão Semântica 2: Emoções e Sociedade

Dimensão Semântica 3: Referência a Música

Se o pesquisador optar por observar características gramaticais e de padronização da língua em geral, as letras de música podem ser analisadas sob as seguintes dimensões:

Dimensão Léxico-gramatical 1: Argumentação e Padronização

Dimensão Léxico-gramatical 2: Interação e Descrição

Dimensão Léxico-gramatical 3: Narrativa no Passado X Referencial imediato

Na seção seguinte, encerramos este capítulo apresentando uma discussão dos resultados em relação aos teóricos que embasaram nossa pesquisa.

### **5.3. Discussão**

Os procedimentos para uma Análise Multidimensional conforme propostos por Biber (1988) são finalizados pela “interpretação de cada fator e rotulação das dimensões” (BERBER SARDINHA, 2004, p. 306), o que fizemos neste capítulo.

A descoberta de dimensões de análise para letras de música relaciona-se especialmente com o caráter exploratório de nossa pesquisa, assim como outras que mencionamos (KAUFFMANN, 2005; PACHECO DE OLIVEIRA, 1995, CROSSLEY; LOUWERSE, 2007), pois não partem de dimensões de análise pré-existentes, mas exploram novas possibilidades.

Os resultados obtidos nesta pesquisa exploram o uso de uma metodologia de cunho estatístico como a AMD para a análise linguística. Assim como os outros autores, ao computarmos os valores de ocorrência lexical por suas características morfossintáticas, pudemos identificar como elas se agrupam e refletem as escolhas dos falantes de acordo com a função que desejam que sua fala ou texto desempenhe. Vale ressaltar aqui que o modelo metodológico da AMD prestou-se para esta análise, comprovando sua validade e replicabilidade, ainda que sejam consideradas características linguísticas diferentes,

dependendo do foco investigativo do pesquisador.

Um aspecto que precisamos destacar em relação à nossa pesquisa quando comparada à dos outros autores por nós mencionados, é a questão da representatividade do corpus, pois, embora nosso objetivo não fosse desvendar dimensões de análise linguística da língua em geral, como no caso da investigação da língua inglesa feita por Biber (1988), o nosso foi o corpus com maior quantidade de palavras (1.151.280) e quantidade de textos (6.290). Por outro lado, Biber (1988) investigou diferentes registros dos gêneros falado e escrito; assim como fizeram Crossley e Louwse (2007), porém com enfoque em pacotes lexicais; Kauffmann (2005), que investigou diferentes subgêneros presentes no jornal e Pacheco de Oliveira (1997), que investigou apenas um gênero, escrita acadêmica de alunos graduandos, mas contrastou representações de língua materna e língua estrangeira.

A seleção de características linguísticas em todas essas pesquisas, assim como na nossa, partiu do mesmo ponto comum (características morfossintáticas) e, de acordo com os objetivos investigativos em cada uma delas, seus pesquisadores selecionaram umas e/ou outras. Todavia, em nossa pesquisa, também consideramos características semânticas e pacotes lexicais.

As dimensões que encontramos também não se apresentam muito distantes das encontradas por esses pesquisadores, considerando-se as diferenças de natureza dos corpora. As dimensões encontradas por Biber (1988), por exemplo, apresentaram características de interação, informação e persuasão, como a nossa. Aliás, a primeira dimensão apresentada por Biber (discurso com interação e envolvimento versus informacional e editado) muito se assemelha à nossa segunda dimensão gramatical (interação versus descrição), da mesma forma que se assemelha à segunda dimensão encontrada por Kauffman (2005) para textos jornalísticos (argumentativo versus informativo) e a segunda dimensão detectada por Pacheco de Oliveira (1997) (orientação discursiva interacional versus informacional). Isso é uma indicação de que, essas funções são desempenhadas pelos falantes em diferentes esferas da linguagem.

A constatação da afinidade de análise multidimensional entre diferentes gêneros nos remete à questão das letras de música enquanto gêneros textuais. Longe do intuito de caracterizar tal gênero estruturalmente, ou pelas características que apresentam



semelhanças à poesia, nossa análise dimensional mostra mais uma semelhança desse tipo de registro com a língua inglesa em geral, em que subjaz a ideia de coloquialismo conversacional presente nas letras de música de acordo com Frith (1998) e, conseqüentemente, por se tratar de música popular de massa, podemos considerar as letras de música tanto como espelho quanto como reflexo da linguagem de um grupo que compartilha certa identidade cultural e social.

Feitas essas considerações, encerramos este capítulo. No capítulo seguinte, apresentamos um resumo dos resultados em relação às nossas perguntas de pesquisa, apresentando nossos comentários finais, inclusive sobre suas limitações e perspectivas.

## 6. Conclusão

Este capítulo encerra a tese. Sendo assim, apresentaremos nossos comentários em vista das perguntas de pesquisa e de nossos resultados. Antes, porém, faremos uma breve revisão sobre nossa motivação e nossos objetivos, das teorias que apoiaram a pesquisa e das decisões metodológicas por nós tomadas. No final do capítulo, serão colocadas algumas críticas e limitações da pesquisa, assim como suas contribuições.

A pesquisa que apresentamos aqui é fruto de nossa motivação pessoal e profissional em relação ao uso de letras de música em sala de aula. Reconhecemos o potencial motivador e pedagógico da música para o ensino de Inglês como língua estrangeira, ao mesmo tempo em que sentimos falta de suporte teórico para um uso mais fundamentado das letras de música no ensino de padrões linguísticos, sejam de ordem lexicais ou gramaticais. Assim, partimos para uma investigação de características linguísticas presentes em letras de música, para que, em nossa prática pedagógica futura, tivéssemos o suporte almejado.

Dois pressupostos principais embasaram toda a pesquisa, desde sua concepção. O primeiro é a visão de língua como um sistema probabilístico, seguindo as ideias de Firth (FIRTH, 1957; HALLIDAY, 1991; SINCLAIR, 1991; HALLIDAY; HASAN, 1989; HALLIDAY; WEBSTER, 2002; BERBER SARDINHA, 2004). Em outras palavras, a Linguística de Corpus, teoria principal em nossa pesquisa, ao considerar essa concepção de língua, admite que as várias possibilidades de escolhas e combinações lexicais dentro de um determinado contexto não ocorrem da mesma forma nem com a mesma frequência. Por outro lado, essas mesmas escolhas e combinações definirão suas funções e seu uso. O segundo pressuposto reside na natureza social da música e a influência das palavras usadas na música para a linguagem cotidiana das pessoas, ao mesmo tempo que observa a influência da linguagem das pessoas para as letras de música. Em outras palavras, ao considerarmos a música popular como aquela produzida para a massa e divulgada em massa pelos meios midiáticos, admitimos seu potencial como fonte de conhecimento linguístico (STARR; WATERMAN, 2007; MOORE, 2003; GRIFFITHS, 2003).

O caráter probabilístico da língua leva em consideração a unidade funcional texto/contexto, ou seja, as variações linguísticas ocorrem de acordo com o contexto e a função que o falante quer desempenhar. Uma vez que consideramos as letras de música como texto, fonte de conhecimento linguístico cuja variação está atrelada a um contexto, consideramos também sua característica funcional e, conseqüentemente, seu pertencimento a um gênero. Adotamos, nesta pesquisa, a perspectiva de análise textual para gênero, cuja representação, na forma de registro, vai indicar os propósitos comunicativos e o contexto situacional de textos.

A distribuição das variações linguísticas presentes em um corpus pode ser explorada a partir de sua observação empírica e deve-se levar em consideração diversos níveis de análise. Ou seja, os textos representam diferentes funções comunicativas que não podem ser observadas sob uma única perspectiva, mas multidimensionalmente.

Ao longo do trabalho, portanto, defendemos a tese de que letras de música carregam valores pertinentes para análise linguística e que as variações entre uma música e outra envolvem mais aspectos do que os puramente estilísticos e sonoros, isto é, as letras de música agrupam-se também de acordo com suas características linguísticas.

Assim, o objetivo principal deste trabalho foi o de desvendar as dimensões de variação linguística que permitem uma caracterização da música popular anglo-americana a partir da constituição léxico-gramatical e semântica de suas letras.

Coletamos um corpus de letras de música compostas originalmente em inglês, contendo 6.290 músicas diferentes, gravadas por 32 artistas diferentes entre os anos de 1940 e 2009, os quais representaram 15 diferentes gêneros musicais (conforme classificação da mídia). Reunidas, as letras de música somaram 1.151.280 palavras.

O tamanho do nosso corpus é um fator de destaque nas pesquisas de análise multidimensional, pois está entre os maiores usados em AMD.

A fim de desvendar as dimensões de variação linguísticas, adotamos a proposta metodológica de análise fatorial, AMD, a partir do modelo de Biber (1988). Seguindo os passos dessa metodologia, tivemos os dados do corpus de estudo etiquetados por suas funções morfossintáticas e por seus agrupamentos semânticos. Foram consideradas

também representações de padronização da língua por meio de pacotes lexicais de três palavras (trigramas), que, foram associados às características morfossintáticas para a extração fatorial. Em seguida, foram feitos diversos cálculos que nos levaram à extração final de três fatores para cada uma dessas esferas de análise. A interpretação desses fatores revelaram as dimensões apresentadas no capítulo anterior. Os resultados obtidos forneceram subsídios para que respondêssemos nossas questões de pesquisa, o que apresentamos pontualmente a seguir.

### **Primeira questão de pesquisa:**

Quais os principais fatores responsáveis pela variação linguística presente nas letras de música popular anglo-americana?

Nossa extração fatorial apontou para a existência de três fatores, de acordo com a perspectiva de análise, como descrito em detalhes no capítulo 5. Quando observadas características gramaticais, os fatores apontam para as seguintes oposições polares: Fator 1 – infinitivos/gerúndios e modais X substantivos; Fator 2 – pronomes pessoais e possessivos X qualificadores; e Fator 3 – verbos no tempo passado X verbos no tempo presente.

Quando observadas as características semânticas, os fatores apontam a presença simultânea dos seguintes agrupamentos: Fator 1 – marcadores de movimento/tempo/fala/pessoas/objetos; Fator 2 – marcadores de emoção e de manifestações sociais; e Fator 3 – marcadores de manifestação musical. Essa composição indica que dentre todas as categorias gramaticais da língua inglesa (97), apenas 13 de fato fazem diferença na distinção entre as letras. Os compositores tem à sua disposição todo o repertório da língua inglesa, mas apenas uma pequena parte dele é discriminatório, quer dizer, pode significar que uma letra de música se pareça mais ou menos com um certo tipo de música. Em termos simples, podemos dizer que a gramática das letras de música é restrita, sendo composta por categorias nucleares que distinguem canção de canção e tornam certas canções mais parecidas com algumas do que com outras. Essa classificação não corresponde à classificação mercadológica conhecida. Não cremos que a classificação mercadológica possa ser substituída pela nossa, visto que a que descobrimos é de caráter mais técnico e verbal, ao passo que a mercadológica leva em conta em essência a melodia, o desenvolvimento dos ritmos ao

longo do tempo e interesses econômicos. Não cremos tampouco que nossa classificação seja estanque – à medida que mais músicas sejam criadas, talvez a classificação resultante mude. Nosso corpus é bastante extenso e variado para nos permitir acreditar que os resultados sejam robustos e não mudem radicalmente com a mudança de corpus, a não ser um que não seja comparável ao nosso, que tenha composição e extensão muito diferentes. Contudo, conforme mostram o próprio Biber (antes de 1988) e Lee (2000), a mudança de corpus pode implicar a alteração nos fatores e nas dimensões.

Quando observados os pacotes lexicais (trigramas) mais frequentes em letras de música e no corpus do Google N-Gram associados às características gramaticais, os fatores apontaram para as seguintes presenças: Fator 1 – infinitivos/gerúndios/modais/pacotes lexicais X substantivos; Fator 2 – pronomes pessoais e possessivos X qualificadores; e Fator 3 – verbos no tempo passado X verbos no tempo presente.

Como vemos, o terceiro conjunto de fatores é uma extensão do primeiro. Eles foram extraídos e analisados, a despeito de suas semelhanças, a fim de se oferecer à futuras pesquisas a opção de se considerarem também os pacotes lexicais que refletem a padronização da língua. Nossa análise também mostrou, assim como para a gramática do inglês, que apenas um conjunto menor de temas discrimina as canções em um tipo ou outro. Teoricamente, os compositores tem à sua disposição qualquer assunto na criação de suas letras, porém o que vemos, na prática, é que apenas um subconjunto dele (8) é de fato relevante para fins de distinguir uma letra da outra. Não temos dados diretos a respeito disso, mas até onde sabemos, os compositores não parecem ter consciência das categorias mais discriminantes da composição musical. Ainda assim, embora escrevam letras sobre diversos temas, em geral, acabam privilegiando determinados assuntos.

### **Segunda questão de pesquisa:**

Quais dimensões (funções comunicativas) expressam os fatores obtidos?

As funções comunicativas por nós identificadas também variam de acordo com a perspectiva de análise. Gramaticalmente observamos as dimensões ARGUMENTAÇÃO X INFORMAÇÃO; INTERAÇÃO X DESCRIÇÃO e NARRATIVA NO PASSADO X REFERENCIAL IMEDIATO. Quando observamos

características semânticas, identificamos as seguintes dimensões: letras de música que falam de AÇÕES PESSOAIS, de EMOÇÕES E SOCIEDADE e letras de música que apresentam REFERÊNCIA A MÚSICA. As dimensões que emergiram da interpretação dos fatores léxico-gramaticais (morfo-sintáticos e trigramas) são: ARGUMENTAÇÃO E PADRONIZAÇÃO ; INTERAÇÃO X DESCRIÇÃO e NARRATIVA NO PASSADO X REFERENCIAL IMEDIATO.

As outras questões de pesquisa envolvem a representação das dimensões pelos dados. Como se trata de uma análise multidimensional, não podemos dizer que determinado gênero/estilo musical, artista ou período não representa, em nenhuma de suas letras, determinada dimensão. Por outro lado, a análise descritiva apontou para aqueles mais e menos representativos em cada dimensão e como os outros representantes se aproximam ou se distanciam desses polos, como descrevemos detalhadamente no capítulo anterior.

A diferença dimensional que detectamos para letras de música parece-nos ocorrer de forma inconsciente, fugindo à percepção de seus autores. Ou seja, seus autores escolhem palavras, estruturas gramaticais e temas intuitivamente e, a intuição, com afirma Sinclair (1991, p.112) não consegue mostrar com clareza o que acontece com a linguagem em uso. Somente uma pesquisa de larga escala, multifatorial, multidimensional, com corpus, poderia levantar esse aspecto. É provável que a maioria dos ouvintes de música também não percebam essas dimensões ou nem se importem, ainda que, algumas vezes consigam perceber alguns temas. Mas não tem dados para comprovar isso. Talvez essa classificação seja abstrata em demasia para ser percebida conscientemente por quem ouve ou escreve música.

Assim, apresentamos a seguir as outras três questões de pesquisa, seguidas por uma tabela contendo os gêneros, artistas e períodos mais representativos de cada dimensão. Ressaltamos que, dada a semelhança entre a segunda e a terceira dimensões gramatical e léxico-gramatical, elas estão representadas na tabela apenas uma vez.

### **Terceira questão de pesquisa:**

Como essas dimensões são representadas em relação aos diferentes gêneros musicais presentes em nosso corpus de estudo?

**Quarta questão de pesquisa:**

Como essas dimensões são representadas em relação aos diferentes artistas selecionados para o estudo?

**Quinta questão de pesquisa:**

Como essas dimensões são representadas ao longo do tempo por nós estudado?

O quadro a seguir responde a essas três perguntas:

Dimensão	maior representatividade		
	gênero	artista	período
<b>Argumentação X Informação</b>	Punk Pop	Simple Plan	anos 2000
<b>Interação X Descrição</b>	Rock'n'roll	Madonna	anos 40
<b>Narrativa Passada X Referencial Imediato</b>	country	Johnny Cash	anos 70
<b>Ações Pessoais</b>	Rock surfista	Beach Boys	1960
<b>Emoções e Sociedade</b>	Heavy Metal	Metallica	1940
<b>Referência a Música</b>	pop vocal	Frank Sinatra	1940
<b>Argumentação e Padronização</b>	Punk Pop	Simple Plan	anos 2000

Quadro 51: Representatividade nas dimensões de análise

A metodologia escolhida neste trabalho, a AMD, não deve ser considerada como única possibilidade de análise linguística para letras de música, mas serviu como modelo para esta pesquisa, oportunizando o contraste de elementos linguísticos em termos funcionais. Pudemos observar que quando consideramos as músicas por suas características linguísticas, nem sempre elas se agrupam da mesma forma que as categorizações genéricas impostas pelos meios de comunicação de massa os quais têm outros objetivos (como entretenimento e comércio), que não os de oferecer uma mensagem ao ouvinte por meio de uma linguagem que lhe faça sentido.

Como toda pesquisa, esta também apresenta limitações. A primeira crítica que fazemos é em relação à composição do corpus. É possível que, se o corpus fosse composto por outras letras de música, por estilos mais variados, ou ainda, com maior

quantidade de palavras, os resultados tivessem sido outros. Contudo, consideramos que uma amostra com quase 1 milhão e duzentas mil palavras possa ser considerada bastante representativa do universo da música popular anglo-americana.

Apresentamos críticas também à metodologia de análise multidimensional. Primeiramente, o trabalho com um programa de análise estatística, o SPSS, foi bastante difícil, mesmo contando com uma interface amigável e explicações detalhadas expressas em outras pesquisas, como nas de Kauffmann (2005) e Pacheco de Oliveira (1997). A análise numérica não é, necessariamente, uma área que possa ser considerada confortável para maioria dos linguistas. Em segundo lugar, a etiquetagem do corpus não pode ser conferida manualmente, dada a dimensão dos dados, o que pode ter causado alguns erros. Por outro lado, acreditamos que a grande quantidade de dados deve também tê-los dissipado ao longo dos textos. Finalmente, uma característica da análise multidimensional que criticamos é a escolha subjetiva dos traços linguísticos que integram a análise fatorial, pois fica a cargo do pesquisador. No caso desta pesquisa, procuramos considerar os mais relevantes, ainda assim, tivemos que agrupá-los, pois apresentaram-se de forma sobreposta, o que dificultava a observação das divisões fatoriais.

Uma limitação de nosso estudo é não termos considerado as possibilidades de prosódia semântica, ou metafóricas, uma vez que acreditamos que a linguagem musical deva estar repleta delas. Essa limitação fica como sugestão para pesquisas futuras, desta pesquisadora ou de outros.

Acreditamos que esta pesquisa oferece bastante contribuição tanto para a área de AMD como para a linguística como um todo. Se por um lado nossa pesquisa seguiu uma metodologia que não era inédita, ela usou de um corpus inédito. Nem mesmo as grandes coleções de exemplos da língua geral contêm letras de música. Provamos neste trabalho que as letras de música são bastante representativas de uma língua e acreditamos que deva ser inserida como modalidade textual nos grandes corpora como BNC, ANC, COCA e Google N-Gram.

Uma característica inovadora de nossa pesquisa foi a seleção de componentes



semânticos para a análise, o que não encontramos na literatura que consultamos. Essa perspectiva de análise gerou três dimensões que exprimem os assuntos mais frequentes das letras de música em nosso corpus: músicas que falam das pessoas, do que elas fazem, quando e com quem; músicas que falam de emoções e de problemas e crenças sociais; e músicas que falam de música e dança.

Nossa pesquisa também pode servir de apoio para o desenvolvimento de atividades pedagógicas e material didático. Professores podem organizar seu material de ensino em cima das dimensões gramaticais, semânticas ou léxico-gramaticais, escolhendo músicas típicas de cada dimensão. Podem, também, selecionar as letras de músicas a partir de características linguísticas determinadas que pretendem focar. Esse levantamento (das características e das dimensões) já está pronto. Com isso, o uso da música pode passar a ser mais efetivo, porque as músicas escolhidas terão conteúdo lingüístico específico, que pode ser explorado diretamente na aula, em si ou em conjunto com o material do curso. Dessa forma, a música pode deixar de ser apenas um momento de descontração na aula, mas um momento de aprendizagem que ocorre com diversão e descontração e, provavelmente, de forma mais efetiva e significativa, vez que os aprendizes podem continuar a ouvir a música fora do ambiente pedagógico.

Acreditamos que nossa pesquisa possa contribuir também para estudos da área de musicologia, que podem considerar dimensões de análise linguística associadas aos fatores de musicalidade e composição, e para a área de ensino de línguas que podem abordar as características lexicais, semânticas, gramaticais e funcionais das letras de música da mesma forma que consideram outras manifestações textuais.

Além disso, não só os estudiosos de música, mas as pessoas envolvidas com o processo de criação e divulgação musical, como as gravadoras e os compositores, podem começar a ‘perceber’ as letras de música dentro das ‘novas’ dimensões que apontamos e, dessa forma, compreender que as categorias lingüísticas que distinguem as músicas são poucas e determinadas. Ou seja, embora esses resultados não tenham aplicação direta na composição musical, as dimensões podem servir de base para iniciar uma discussão e *feedback* de como podem, compositores e amantes da música em geral, descobrir aplicações na musicologia, na criação, na divulgação e na apreciação da música.

Contudo, mais pesquisas precisam ser feitas sobre isso.

Outro ponto importante é que nossas dimensões mostraram que os compositores, mesmo tão díspares e distantes cronologicamente, usam a linguagem de modo parecido, de tal modo que suas canções formam agrupamentos atemporais, muitas vezes. É interessante ver que mesmo músicas de épocas tão distantes se assemelham, sem falar de músicas de estilos que supostamente deveriam ser muito diferentes de si, como algumas músicas gravadas por Frank Sinatra nos anos 40 e as músicas gravadas por bandas de heavy metal nos anos 90, como Iron Maiden. Isso pode significar que há uma ‘linguagem verbal da música’ que os compositores conhecem, talvez instintivamente ou por experiência no ofício, talvez instigada pela exposição à música e pelo gosto por ela.

Finalmente, para os amantes de música, em especial aqueles que se consideram como ecléticos, ou seja, apreciam diferentes estilos musicais, considerados incompatíveis por muitas pessoas, descobrimos um ponto de intersecção que esperamos que possa esclarecer e justificar linguisticamente seu gosto. Pode ser que, a despeito da sonoridade, essas pessoas gostem mesmo daquilo que a música fala e como ela fala.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ARLEO, A. (2000). Music, song and foreign language teaching. **Les Cahiers de l'APLUIT**, XIX, (4): 5-19.
- BAKHTIN, M. Os gêneros do discurso. In: **Estética da criação verbal**. São Paulo: Martins Fontes, (1953), 1992.
- BENNETT, Andy. The forgotten decade: rethinking the popular music of the 1970's. **Popular Music History** 2 (1), 2007 (p.5-24)
- BERBER SARDINHA, A. P. **Linguística de Corpus**. Barueri: Manole, 2004 (a).
- \_\_\_\_\_. Linguística de Corpus: histórico e problemática. **D.E.L.T.A.**, São Paulo, n. 16, v.2, p. 323-367, 2000 (a)
- \_\_\_\_\_. Análise Multidimensional. **D.E.L.T.A.**, São Paulo, n. 16, v.1, p. 99-127, 2000 (b)
- \_\_\_\_\_. Informatividade, interatividade e narratividade na reunião de negócios – Análise Multidimensional e palavras-chave. **DIRECT Papers**, São Paulo e Liverpool, n. 52, 2004 (b).
- \_\_\_\_\_. Análise de gênero e Linguística de Corpus: identificação das unidades internas do gênero por meio da padronização lexical. **DIRECT Papers**, São Paulo e Liverpool, n. 51, 2003.
- \_\_\_\_\_. **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005.
- \_\_\_\_\_. **Metáfora**. São Paulo: Parábola, 2007.
- BERLAND, J. Sound, image and social space: music vídeo and media reconstruction. In: FRITH, S., A. GOODWIN; L. GROSSBERG (eds.), **Sound and vision: the music video reader**. London, UK: Routledge, 2003. (1993 first print). Cap. 2 (25-43)
- BÉRTOLI-DUTRA, P. **Explorando a linguística de corpus e letras de música na produção de atividades pedagógicas**. Dissertação de Mestrado Inédita, LAEL, PUC-SP, 2002
- BIBER, D. **Variation across speech and writing**. Cambridge: Cambridge University Press, 1988

- \_\_\_\_\_. **Dimensions of Register Variation: A Cross Linguistic Comparison.** Cambridge: Cambridge University Press, 1995.
- \_\_\_\_\_. **University Language.** A corpus based study of spoken and written register. Amsterdam/Philadelphia: John Benjamins, 2006.
- BIBER, D.; S. CONRAD. **Register, Genre, and Style.** Cambridge: Cambridge University Press, 2009.
- BIBER, D.; S. CONRAD; R. REPPEN. **Corpus Linguistics: Investigating Language Structure and Use.** Cambridge: Cambridge University Press, 1998.
- BIBER, D.; S. CONRAD; V. CORTES. If you look at ... lexical bundles in academic lectures and textbooks. **Applied Linguistics**, Oxford, v. 25, n. 3, p. 371–405, 2004.
- BRACKETT, David. **Interpreting Popular Music.** University of California Press, 2000.
- BRANTS, T.; A. FRANZ. Web 1 T5-gram Version 1. Linguistic Data Consortium: Philadelphia. 2006. Disponível em < <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13>>
- CHARLTON, K. **Rock music styles: A History.** Boston: MacGraw-Hill, 2008.
- COSTA, Nelson Barros da. **A produção do discurso lítero-musical brasileiro.** Tese de doutorado apresentada ao Programa de Estudos Pós-graduados de Linguística Aplicada e Estudos da Linguagem (LAEL), PUC-SP, 2001.
- \_\_\_\_\_. As letras e a letra: o gênero canção na mídia literária. In: DIONÍSIO, A.P.; A. R., MACHADO; M. A. BEZERRA. (orgs.). **Gêneros textuais e ensino.** Rio de Janeiro: Lucerna, 2002.
- CROSSLEY, S.A. Metaphorical conceptions in hip-hop music. **African American Review**, 9 (4) 2005. p. 501-512. Disponível em <[http://www.msstate.edu/dept/english/faculty/crossleypubs/Metaphorical\\_conceptions\\_in\\_hip-hop\\_music.pdf](http://www.msstate.edu/dept/english/faculty/crossleypubs/Metaphorical_conceptions_in_hip-hop_music.pdf)>. Acesso em: 18 jun. 2008.
- CROSSLEY, S.A. ; M., LOUWERSE. Multi-dimensional register classification using bigrams. **International Journal of Corpus Linguistics**, 12 (4): 453-478, 2007.

- DAMAZO, F.A.F.T. **O canto do povo de um lugar: uma leitura das canções de João do Vale**. Tese de doutorado apresentada ao Programa de Estudos Literários da Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP: SJ Rio Preto, 2004
- DANTAS, D. F. **A prateleira do rock brasileiro: uma análise das estratégias midiáticas utilizadas nos discos de rock brasileiro nas últimas cinco décadas**. Dissertação de mestrado apresentada ao Programa de Pós Graduação em Comunicação e Cultura Contemporâneas de Comunicação da Faculdade de Comunicação da Universidade Federal da Bahia, Salvador, 2007.
- EGGINS, S; J. R. MARTIN. Genres and registers of discourse. In: T.A. VAN DICK (ed.). **Discourse as structure and process**. Sage Publications, 1997.
- FERREIRA, A.B.H. **Novo Dicionário da língua portuguesa**. 3.ed. Rio de Janeiro: Nova Fronteira, 1999.
- FIRTH, Alan. **The discursive accomplishment of normality on ‘língua franca’ English and conversation analysis**. Journal of Pragmatics 26. 1996 (237-259)
- \_\_\_\_\_. **Papers in Linguistics – 1934-1951**. Oxford: Oxford University Press, 1957.
- FREDERICK, J. **Up from trigrams!** The struggle for improved language models. In Proceedings of Eurospeech-91, volume 3, pages 1037–1040, Genova, Italy, 1991
- FRITH, S. Music and identity. In: HALL, Stuart; Du GAY, Paul (Eds). **Questions of Cultural Identity**, London, UK: Sage publications, 1993. p. 108-127
- \_\_\_\_\_. **Performin Rites**. On the value of popular music. Cambridge, Massachusetts, USA: Harvard University Press, 1998.
- FRITH, S., A. GOODWIN; L. GROSSBERG. (eds.), **Sound and vision: the music video reader**. London, UK: Routledge, 2003.
- GEORGE, N. **The Death of Rhythm and Blues**. New York: Peguin Books, 1988.
- GILLON, Steven. **Boomer Nation**. The largest and riches generation ever and how it changed America. New York: Free Press, 2004.

- GOODWIN, A. MTV meets postmodernity theory. In: FRITH, S, A. GOODWIN; L. GROSSBERG. (eds.), **Sound and vision: the music video reader**. London, UK: Routledge, 2003. (1993 first print). Cap. 3 (45-66)
- GRIFFITHS, Dai. From Lyrics to anti-lyric: analyzing the words in pop songs. In: A. F. MOORE. (ed). **Analysing Popular Music**. Cambridge: Cambridge University Press, 2003. (p. 39-59)
- HALL, S. **A identidade cultural na pós-modernidade**. 11<sup>a</sup> ed. Rio de Janeiro: DP&A, 2006.
- HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. In: K. Aijmer; B. Altenberg (org.) **English Corpus Linguistics: Studies in honour of Jan Svartvik**. (30-43). London: Longman, 1991.
- \_\_\_\_\_. **Introduction to a Functional Grammar**. Second Edition. London: Edward Arnold, 1994.
- \_\_\_\_\_. Quantitative studies and probabilities in grammar. In M. Hoey (ed.). **Data, description, discourse**. Harper Collins, 1993.
- HALLIDAY, M.A.K; R. HASAN. **Language, context, and text: aspects of language in a social-semiotic perspective**. 2nd edition. Deakin University Press/Oxford University Press, 1989.
- HALLIDAY, M.A.K; J. WEBSTER (ed.). **On grammar: By Michael Alexander Kirkwood Halliday**. New York: Continuum, 2002.
- HUNSTON, S. **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press, 2002.
- HUNSTON, S.; FRANCIS, G. **Pattern grammar: a corpus-driven approach to the lexical grammar of English**. Amsterdam/Philadelphia: John Benjamins, 2000.
- JANOTTI JR, J. S. **Heavy Metal: o universo tribal e o espaço dos sonhos**. Dissertação de mestrado apresentada ao Programa Pós-graduação em multimeios da UNICAMP, 1994.
- KAUFFMANN, C. H. **Corpus do jornal: variação linguística, gêneros e dimensões da imprensa diária escrita**. Dissertação de mestrado apresentada ao Programa de

Estudos Pós-graduados de Linguística Aplicada e Estudos da Linguagem (LAEL), PUC-SP, 2005.

KENNEDY, G. D. **An Introduction to Corpus Linguistics**. London: Longman, 1998.

KRIEGER, D. Corpus Linguistics: What it is and how it can be applied to teaching. **The Internet TESL Journal**. 2003. Disponível em: <<http://iteslj.org/Articles/Krieger-Corpus.html>>

LAFFERTY, J.; D. SLEATOR; D. TEMPERLEY, **Grammatical trigrams: A probabilistic model of link grammar**, in Proceedings of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language, Cambridge, MA, October 1992.

LEE, D.Y.W. **Modelling variation in spoken and written language: the Multi-Dimensional Approach revisited**. Lancaster, 2000. Tese (Doutoramento em Lingüística) \_ Dept. Of Linguistics and Modern English Language, Lancaster University.

MATTE, A.C.F. Gostar de música: percurso de uma paixão. **Significação: revista brasileira de semiótica**. 23. São Paulo: ANNABLUME. Junho, 2005. p. 72-92

McENERY, T.; A. WILSON, J. **Corpus Linguistics**. Edinburgh: Edinburgh University Press, 1996.

MIDDLETON, Richard (ed). **Reading pop**. Approaches to textual analysis in popular music. Oxford University Press, 2000.

\_\_\_\_\_. **Studying popular music**. Open University Press, 1990.

MOORE, Allan F. (ed). **Analysing Popular Music**. Cambridge: Cambridge University Press, 2003.

MORA, C. F. Foreign language acquisition and melody singing. **ELT Journal**, 54 (2):146-152, 2000.

MUNDY, John. **Popular music on screen: from hollywood musical to music video**. Manchester and New York: Manchester University Press, 1999.

MURPHEY, T. The song stuck in my head phenomenon: a melodic din in the lad? **System**, 18 (1): 53-64, 1990.

- NESTROVSKI, A. **Notas Musicais: Do Barroco ao Jazz**. São Paulo: Publifolha, 2000.
- O'KEEFE, A.; M. MacCARTHY; R. CARTER. **From corpus to Classroom: language use and language teaching**. Cambridge: Cambridge University Press, 2007.
- PACHECO DE OLIVEIRA, L. **Variação intercultural na escrita: contrastes multidimensionais em inglês e português**. Tese de doutorado apresentada ao Programa de Estudos Pós-graduados de Linguística Aplicada e Estudos da Linguagem (LAEL), PUC-SP, 1997.
- PEDERSEN, T., **Unsupervised Corpus Based Methods for WSD**. In Agirre, E. and Edmonds, P. (Editors), **Word Sense Disambiguation : Algorithms and Applications**, June 2006, pp. 133-166.
- PENN Tree-Tagger: consultar em: <http://www.cis.upenn.edu/~treebank/> e <http://search.cpan.org/~tpederse/WordNet-SenseRelate-AllWords-0.19/Utils/wsd.pl>
- RAMILO, M.C.; T. FREITAS. **A linguística e a linguagem dos média em Portugal: descrição do projeto REDIP**. In XIII Congresso Internacional da ALFAL. San José, Costa Rica, 2002.
- ROSSATO, Ednéia Vieira . **A brasilidade na poesia contemporânea de Caetano Veloso**. In: III SEMINÁRIO DE ESTUDOS SOBRE LINGUAGEM E SIGNIFICAÇÃO - SELISIGNO: DISCURSO E REPRESENTAÇÃO E IV SIMPÓSIO DE LEITURA DA UEL, 2002, Londrina. Caderno de Resumos do IV Simpósio de Leitura da UEL: III Selisigno: Discurso e Representação, 2002. v. 1. p. 93-93.
- SAMPSON, G. **Empirical Linguistics**. London: Continuum, 2001.
- SANTOS, V.B.M.P. Estabelecendo as diferenças entre os termos registro e gênero. **DIRECT Papers**, São Paulo, n. 33, 1996.
- SCHOEPP, K. Reasons for using songs in the ESL/EFL classroom. **The Internet TESL Journal**, vol VII, No. 2, February, 2001. Disponível em <<http://iteslj.org/Articles/Schoepp-Songs.html>> Acesso em: 02 jun. 2007.
- SHUKER, ROY. **Understanding popular music**. London, New York: Routledge, 1994.
- \_\_\_\_\_. **Popular Music: Key concepts**. London, New York: Routledge, 2005.



- SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.
- SINCLAIR, J. M.; R. Carter. **Trust the text: Language, corpus and discourse**. Oxford: Routledge, 2004.
- STARR, L.; C.WATERMAN. **American Popular Music**. From minstrelsy to MP3. 2nd. ed. New York: OUP, 2007
- STRAW, W. Pop music and postmodernism in the 1980s. In: FRITH, A. GOODWIN; L. GROSSBERG. (eds.), **Sound and vision: the music video reader**. London, UK: Routledge, 2003. (1993first print). Cap. 1 (3-21)
- STUESSY, J. **Rock and Roll: Its history and stylistic development**. 6th. Ed. New York: Prentice Hall, 2008
- SWALES, J.M. **Genre Analysis** – English in academic and research settings. Cambridge: Cambridge University Press, 1990.
- SZATMARY, D. P. **Rockin' in time**. A social history of rock-and-roll. 4th. ed. New Jersey, USA: Prentice Hall, 2000.
- TATIT, L. **O cancionista: composição de canções no Brasil**. São Paulo: EDUSP, 1996.
- \_\_\_\_\_. **Análise Semiótica através das letras**. São Paulo: Ateliê Editorial, 2001.
- THOMPSON, B; S. A. VIDAL-BROWN, **Principal Components versus Principal Axis Factors: when will we ever learn?** Paper apresentado no Annual Meeting of the Southwest Educational Research Association. New Orleans, LA, 2001
- TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam e Philadelphia: John Benjamins, 2001.
- WINKLER, Peter. Randy Newman's Americana. In: R. MIDDLETON (ed). **Reading pop**. Approaches to textual analysis in popular music. Oxford University Press, 2000. (p.27-57)
- WISNIK, J. M. **O som e o sentido**. Uma outra história das músicas. 2<sup>a</sup>. ed. São Paulo: Companhia das Letras, 1989.

# ANEXOS

## ANEXO 1

## LISTA DOS 200 TRIGRAMAS MAIS FREQUENTES DO CORPUS DE ESTUDO

A seguir, apresentamos tabela contendo a lista dos 200 trigrams mais frequentes em nosso corpus de estudo, precedidos por sua frequência.

630 i love you	154 why don't you	117 i've got a	98 i feel so
550 i want to	154 the end of	116 got to be	98 ha ha ha
499 oh oh oh	154 me to the	115 the same old	97 you you you
426 yeah yeah yeah	153 you love me	115 just a little	97 rock and roll
382 i don't know	153 you don't know	115 a lot of	97 and i will
303 hey hey hey	153 want you to	113 me in the	96 take my hand
279 da da da	153 i don't need	112 on the ground	96 i wish i
275 i got a	152 the only one	112 love love love	96 be the one
267 la la la	151 what i want	112 for you to	96 and you know
259 i don't want	151 i need you	112 do do do	95 you give me
249 na na na	150 you know you	111 in the dark	95 out in the
248 i know that	150 it's time to	109 know what i	95 look at me
246 no no no	150 and i don't	108 the other side	95 i tried to
245 don't want to	148 you make me	108 don't need no	95 for you and
232 in love with	147 all the way	107 i just want	95 duh duh duh
227 you and i	146 oo oo oo	107 but i don't	94 you don't have
226 i want you	145 as long as	107 and i can't	94 so hard to
216 you want to	144 night and day	106 me and i	94 my love is
213 you know i	141 used to be	106 in the air	94 i know now
209 i don't wanna	138 don't know what	105 you want me	94 don't know why
203 on and on	136 i used to	105 you have to	94 do you know
200 i just can't	136 i can see	105 i feel like	94 a long long
197 you and me	135 tell me what	105 i am the	93 it's all right
197 in my heart	134 you know what	105 don't have to	93 down to the
195 want to be	134 if you don't	104 there is no	93 all night long
195 in your eyes	134 i'll be there	104 of my heart	93 ah ah ah
189 come on come	134 fall in love	104 know that i	92 your eyes and
187 if you want	132 love with you	104 i got the	92 you know the
185 on come on	132 i've got to	104 baby baby baby	92 to tell you
185 and if you	132 all i want	103 what do you	92 i was a
183 i know i	130 i know you	103 i got to	91 little bit of
183 a little bit	129 what you want	102 my heart and	91 give me a
182 can't you see	128 in the morning	102 i'd like to	90 my heart is
181 in the sky	127 the way i	102 back to the	90 love you i
180 to be a	127 take me to	102 all the things	90 it's just a
176 now i know	126 but i know	101 you've got to	90 i tell you
172 out of my	125 you tell me	101 let it be	90 i had to
171 if i could	125 let me go	101 i need to	90 but i can't
171 i don't care	123 to me i	101 come on baby	89 you let me
171 don't you know	122 of my life	100 you know i'm	89 what to do
169 you know that	122 i have to	100 to me and	89 the way that
168 and i know	122 and all the	100 of the night	89 me if you
167 you got to	121 when i was	100 let me be	89 i miss you
165 make me feel	120 you got me	99 up in the	89 i got my
165 all the time	120 the things that	99 on my mind	89 i can't get
160 the way you	119 back to me	99 me and you	88 through the night
159 in the night	118 to see the	99 i need your	88 the sound of
159 i wanna be	118 one more time	99 and if i	88 on my own
157 this is the	118 i'm in love	99 and i'll be	88 let it go
155 out of the	117 there ain't no	98 to be the	88 i'm just a

## ANEXO 2

### AMOSTRA DA PLANILHA DE ANÁLISE FATORIAL

A seguir, reproduzimos nossa planilha de análise fatorial parcialmente, pois, em sua totalidade, a planilha contém 6.291 linhas e 98 colunas. A imagem a seguir contém as 98 colunas, mas apenas para 30 linhas da planilha.

Veremos que na primeira coluna, *file*, estão dispostos os nomes dos arquivos, ou seja, se cada uma das 6.291 letras de música. Na primeira linha após os títulos, por exemplo temos o seguinte nome de arquivo:

```
1965_rs_cry_to_me.txt.token.tag
```

O número indica o ano de gravação da música (1965). Este é seguido pela indicação da banda que a gravou, “rs” significa *Rolling Stones*. Em seguida há o nome da música, “*cry to me*”. As outras letras referem-se à extensão dos arquivos, do modo como devem ser enviados para o programa de análise fatorial.

Nas próximas colunas estão dispostas as contagens de ocorrência para cada uma das variáveis. No caso da letra da música “*cry to me*”, veremos que contém 208 palavras corridas (ocorrências ou *tokens*); 76 palavras diferentes (*types*); 5 ocorrências de conjunções coordenativas (CC); 1 ocorrência de número cardinal (CD); 8 ocorrências de determinantes; 1 ocorrência de referência existencial (EX); 3 ocorrências de palavras estrangeiras (FW), e assim por diante.

file	tokens	types	CC	CD	DT	EX	FW	IN	JJ	JJR	JJS	LS	MD	NN	NNS
1965 rs cry to me.txt.token.tag	208	76	5	1	8	1	3	22	3	1	0	0	1	53	1
1965 rs down home girl.txt.token.tag	183	103	6	0	7	0	0	18	7	0	0	0	1	38	5
1965 rs down the road a piece.txt.token.tag	165	74	6	2	18	0	0	20	4	0	0	0	4	34	6
1965 rs get of my cloud.txt.token.tag	305	136	6	3	19	1	0	32	8	0	0	0	0	51	8
1965 rs good times.txt.token.tag	207	81	1	2	23	0	14	6	10	2	0	0	8	40	4
1965 rs gotta get away.txt.token.tag	143	64	0	0	7	0	0	6	5	1	0	0	2	11	5
1965 rs heart of stone.txt.token.tag	207	88	3	0	17	0	0	16	8	0	0	0	0	38	2
1965 rs hey crawdaddy.txt.token.tag	189	50	3	12	7	0	0	3	4	0	0	0	0	26	3
1965 rs hitchhike.txt.token.tag	196	73	2	0	13	0	0	11	4	0	0	0	1	63	10
1965 rs little red rooster.txt.token.tag	101	52	2	0	8	0	0	10	14	0	0	0	0	24	4
1965 rs look what youve done.txt.token.tag	132	80	7	1	10	0	0	7	7	0	0	0	1	26	1
1965 rs mercy mercy.txt.token.tag	179	74	5	2	4	0	1	11	3	0	1	0	0	50	3
1965 rs mona.txt.token.tag	179	54	2	0	6	0	0	8	2	0	0	0	6	16	4
1965 rs off the hook.txt.token.tag	152	72	1	0	23	0	0	23	10	0	0	0	0	33	3
1965 rs oh baby we got a good thing.txt.token.tag	197	74	3	0	17	0	3	10	14	0	0	0	1	41	2
1965 rs one more try.txt.token.tag	259	97	5	1	8	0	0	19	3	1	0	0	4	42	9
1965 rs pain in my heart.txt.token.tag	130	48	1	2	5	0	2	10	5	0	0	0	2	27	2
1965 rs she said yeah.txt.token.tag	229	73	1	1	7	0	0	12	9	0	0	0	0	74	3
1965 rs surprise surprise.txt.token.tag	200	75	6	0	1	0	0	13	7	0	0	0	5	18	17
1965 rs talking about you.txt.token.tag	210	81	6	0	11	0	0	7	3	0	0	0	1	41	3
1965 rs thats how strong love is.txt.token.tag	276	73	4	0	10	0	0	8	48	0	0	0	1	42	6
1965 rs the singer not the song.txt.token.tag	123	57	2	0	14	0	0	10	12	0	0	0	3	15	3
1965 rs the under assistant .txt.token.tag	228	104	3	1	18	0	0	14	20	0	0	0	0	25	5
1965 rs what a shame.txt.token.tag	91	48	1	0	11	0	0	4	3	0	0	0	3	17	0
1965 rs you better move on.txt.token.tag	157	80	7	0	6	0	0	11	7	3	0	0	3	11	5
1965 rs you cant catch me.txt.token.tag	260	148	4	3	17	0	0	27	14	0	0	0	1	46	7
1966 bb i just wasnt made for these times.txt.token	261	95	7	0	14	0	0	20	11	0	0	0	8	26	15
1966 bb i know theres an answer.txt.token.tag	143	65	9	0	6	0	0	10	6	0	0	0	7	11	4
1966 bb im waiting for the day.txt.token.tag	212	90	12	2	7	0	0	14	7	0	1	0	8	15	4
1966 bb dont talk.txt.token.tag	141	56	10	0	0	1	0	9	4	0	0	0	4	27	8
1966 bb god only knows.txt.token.tag	183	53	1	0	2	1	0	21	3	0	0	0	9	8	1
1966 bb here today.txt.token.tag	227	106	13	0	10	0	0	13	16	1	0	0	2	33	9
1966 bb sloop john B.txt.token.tag	267	93	6	0	22	0	0	11	7	0	1	0	0	21	2

NP	NPS	PDT	POS	PP	PP	RB	RBR	RBS	RP	SYM	TO	UH	VB	VBD	VBG	VBN	VBP	VBZ	VH	VHD	VHG	VHN	VHP	VHZ	VV	VVD
13	0	0	0	5	21	9	0	0	0	0	14	3	1	0	0	0	0	1	1	0	0	0	0	0	11	2
15	0	0	0	4	19	19	0	0	3	0	4	5	0	2	0	1	0	0	0	0	0	0	0	0	13	1
7	0	0	0	6	11	18	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	15	1
25	0	0	0	19	34	17	0	0	16	0	4	0	0	5	0	2	2	2	0	0	0	0	2	1	8	7
15	0	0	0	5	15	14	0	0	0	0	6	0	2	0	0	0	1	0	0	0	0	0	0	0	16	8
14	0	1	0	6	11	11	0	0	7	0	12	1	0	0	0	0	1	0	0	0	0	0	0	0	17	8
12	0	0	0	2	18	26	2	0	3	0	1	13	0	0	0	1	0	0	0	0	0	0	0	0	18	6
27	0	0	0	0	22	27	0	0	0	0	0	13	0	0	0	1	0	0	0	0	0	0	0	0	16	15
42	0	1	0	4	4	4	0	0	1	0	8	0	1	0	0	0	0	1	0	0	0	0	3	0	6	4
1	0	0	0	3	6	4	0	0	0	0	6	0	0	0	0	1	2	0	0	1	1	0	0	0	7	0
1	0	1	0	7	14	9	0	0	0	0	0	0	0	0	0	0	0	4	0	1	1	0	0	0	8	2
22	0	0	0	4	19	4	0	0	2	0	5	1	0	0	0	1	0	1	0	0	0	0	13	0	8	4
36	0	0	0	4	28	24	0	0	2	0	4	2	0	0	0	0	0	0	0	0	0	0	0	0	14	0
19	0	2	0	5	6	7	0	0	1	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	4	6
12	0	0	0	1	24	26	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	6	12
11	0	2	0	3	37	16	9	0	4	0	6	0	1	0	0	0	0	0	0	0	0	0	0	0	44	7
19	0	0	0	13	9	2	0	0	1	0	0	8	3	0	0	2	1	0	0	0	0	0	0	0	5	2
5	0	0	0	4	31	24	0	0	0	0	8	1	0	0	0	0	0	1	0	0	0	0	0	0	11	11
18	0	1	0	6	39	6	0	0	3	0	3	1	0	13	0	1	0	0	0	1	1	0	2	0	12	10
12	0	2	0	9	42	22	0	0	2	0	7	0	2	1	0	0	0	0	0	0	0	0	0	0	15	2
12	0	0	0	26	15	18	0	0	0	0	2	3	5	1	0	0	0	28	0	0	0	0	0	0	12	1
5	0	0	0	0	18	19	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	0	0	0	2	0
34	0	0	0	4	23	22	0	0	1	0	1	6	1	0	0	0	8	0	0	1	1	0	4	0	9	3
0	0	0	0	1	11	8	0	0	1	0	3	0	3	0	0	0	0	0	0	0	0	0	0	0	6	4
7	0	0	0	6	32	15	1	0	3	0	6	1	1	0	0	1	0	0	0	1	1	0	0	0	14	1
25	1	1	0	6	34	14	0	0	3	0	3	0	1	5	0	0	0	0	0	0	0	0	0	0	16	11
6	0	0	0	6	43	26	0	0	0	0	6	0	1	0	0	1	0	0	0	0	0	0	0	0	14	12
0	0	1	0	4	31	12	2	0	2	0	3	0	2	0	0	0	0	0	0	0	0	0	3	0	12	0
6	0	0	0	4	46	13	0	0	3	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	23	17
4	0	0	0	20	6	15	0	0	0	0	1	0	4	0	1	0	1	0	0	0	0	0	0	0	19	5
27	0	0	0	0	28	25	0	0	0	0	3	0	11	0	0	0	1	0	0	0	0	0	0	0	13	0
9	0	1	0	7	25	28	0	0	0	0	6	0	1	1	0	0	0	6	1	0	0	0	0	0	8	10
25	0	1	0	2	37	42	0	0	5	0	2	0	0	0	0	1	0	1	0	1	1	0	0	0	35	15

WG	WN	WP	WZ	WD	WP	WP	WRE	artifact	body	motio	capability	dancing	day	death	emotio	re	equa	female	group	O	home
13	0	11	1	0	0	0	3	3	1	0	0	0	0	0	0	0	0	7	0	0	6
4	2	7	2	0	0	0	0	0	3	0	0	3	3	0	0	0	0	0	0	0	0
1	3	7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1
2	4	23	4	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
6	0	6	1	0	0	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0
1	6	9	0	0	0	0	1	4	0	0	0	0	0	0	0	1	1	3	0	0	1
4	1	13	1	0	0	0	2	1	3	0	0	0	0	0	0	0	0	0	0	0	0
2	0	6	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	3	0	0	0
9	0	2	0	0	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	1	2
0	2	4	2	0	0	0	0	0	7	0	0	0	7	1	1	0	0	0	0	0	0
3	9	3	2	0	0	7	0	0	1	0	0	0	3	0	0	0	0	2	0	0	1
5	1	6	1	0	0	2	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0
1	0	16	2	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	0	1	0	0	0	1	1	0	0	0	2	0	0	0	0	0	0	0	0	0	3
4	0	8	2	0	0	1	4	0	1	0	0	0	1	0	12	0	0	1	0	0	0
6	0	17	2	1	0	0	1	0	3	0	0	0	4	0	0	0	0	0	0	0	0
0	0	5	1	0	0	0	3	0	7	0	0	0	0	0	0	0	0	1	0	0	0
1	6	17	0	0	0	1	1	0	3	0	0	0	0	0	3	0	0	0	0	0	0
10	3	1	0	0	0	0	3	0	2	0	0	0	0	0	0	0	0	5	0	0	0
6	1	13	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	3	0	0	0	0	27	0	1	0	0	0	0	0	0	0	3	0	0	0	0
3	0	10	1	2	0	0	0	0	5	1	0	0	2	0	0	0	0	0	0	0	0
9	0	6	2	1	0	0	7	0	0	0	0	0	0	0	0	2	0	0	0	0	0
2	0	1	4	0	0	8	0	0	3	0	0	0	2	0	1	0	0	2	0	0	0
3	0	9	0	1	0	2	0	0	2	0	1	0	3	0	0	0	0	0	0	0	1
2	7	12	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
3	8	25	3	0	0	2	4	0	4	0	0	0	7	0	1	0	0	0	0	1	0
0	0	13	0	0	0	3	2	0	0	0	0	0	3	0	0	0	0	0	0	0	0
5	2	10	0	0	0	1	7	0	2	0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	15	0	0	13	0	1	0	0	0	0	9	0	0	0	0	0	0	0	0
2	8	12	5	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	20
0	0	20	3	0	0	0	7	0	0	0	0	0	2	0	0	0	0	1	0	0	1



human	ImpactI	Intentiona	Intent_Psyc	male	motion	music	needs	not	PastFn	process	proposition	pursuing	radiantLi	radiants
1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	1	0	0	2	0	1	0	1	0	4	0	0	0	0
2	0	0	1	0	0	1	0	1	0	2	0	0	0	1
4	0	0	4	0	5	0	0	1	0	3	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	4	0	0	0	0	17	0	0	2	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	1	0	0	0	1	3	0	0	0	0
0	0	0	0	4	0	0	0	0	0	4	0	2	0	4
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	2	0	0	0	0	0	0	1	0	0	0	0	0
1	2	0	0	0	0	0	2	2	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	2	0	0	0	0
0	0	0	4	0	0	0	0	2	1	1	0	0	0	0
0	0	24	1	0	1	0	1	0	0	3	0	0	0	0
2	0	2	0	0	0	0	0	0	0	0	0	0	0	0
1	0	3	0	0	0	0	2	0	0	1	0	1	0	0
0	0	0	0	0	0	2	0	2	0	3	0	0	0	0
7	0	4	5	0	0	0	0	0	0	1	4	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	10	0	6	0	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
0	0	0	1	0	0	0	0	2	0	2	0	0	0	0
0	0	0	0	0	0	0	0	5	0	3	0	0	0	0
1	3	0	0	0	0	1	0	0	0	2	0	0	0	0
4	0	1	1	0	0	0	0	1	0	6	0	0	0	0
0	0	1	0	0	0	0	0	0	1	3	1	0	0	0
1	0	1	2	0	1	0	2	0	0	3	0	0	0	0
0	1	0	0	0	0	0	0	1	0	0	0	0	0	1
1	0	0	1	0	0	0	1	4	0	0	0	0	0	0
0	0	1	1	1	0	0	0	2	0	4	0	0	0	0
1	0	0	0	0	0	0	1	0	0	3	0	0	0	0
1	0	0	0	0	0	0	1	0	0	3	0	0	0	0
0	0	2	3	0	0	1	1	4	0	1	0	0	0	0



requestir	ing in social r	speaking	stating	Subjectivetime	lnitime	m transpo	TRUE	unhapin	walking	wants	fantasti	drugs	religion
0	0	0	1	6	0	0	2	0	0	0	0	1	11
1	0	0	0	5	0	0	0	2	0	0	0	0	7
0	0	7	0	17	0	0	2	0	0	0	0	0	2
0	0	1	0	17	1	6	0	0	0	0	1	0	5
0	0	0	0	2	0	0	0	3	0	1	0	0	8
0	0	0	0	9	0	0	1	1	2	1	0	0	1
2	0	11	0	23	0	2	1	0	0	0	1	0	25
1	0	1	0	1	0	0	2	0	29	0	0	0	2
0	0	0	0	9	0	0	0	0	1	0	0	0	1
1	0	1	0	12	0	1	2	0	0	0	0	2	6
0	0	1	1	16	0	0	1	0	0	0	0	0	12
0	0	8	0	16	0	0	2	0	0	2	0	0	2
1	0	1	0	6	0	2	0	0	0	0	0	0	0
0	0	2	0	17	0	0	2	0	0	2	0	0	5
1	0	0	0	11	0	0	0	4	0	1	0	0	1
0	0	5	0	6	0	0	0	0	10	0	3	0	7
0	0	6	2	31	0	0	0	0	0	0	7	0	8
0	0	1	1	14	0	0	0	0	0	1	0	2	2
1	0	1	0	14	0	3	1	1	1	0	1	0	3
0	0	0	1	39	0	0	3	0	0	23	0	0	1
0	0	0	0	11	0	0	1	0	0	0	0	0	3
0	0	1	0	10	0	0	4	4	0	0	0	1	7
0	0	0	2	16	0	0	1	0	0	0	0	0	4
3	0	2	3	13	0	0	2	0	0	5	1	0	1
0	1	0	0	14	0	0	3	0	0	1	0	0	9
0	0	1	0	17	0	10	3	0	0	1	1	0	8
0	0	0	2	7	0	0	0	0	0	0	2	2	6
0	0	1	0	13	0	0	4	0	1	5	3	1	2
0	0	6	0	5	1	0	0	0	0	0	0	0	5
0	0	1	0	6	2	0	2	0	0	1	0	0	1
0	0	0	0	23	0	1	10	1	3	6	0	0	20
3	0	1	0	15	0	0	20	0	0	0	0	0	9
0	0	2	0	17	0	2	2	0	0	1	0	2	3

sex	violencesong	bundlegoogle	count
3	0	16	40
7	0	2	27
3	0	1	27
4	0	3	88
2	0	14	44
2	0	5	36
16	0	3	28
0	0	23	36
0	0	5	26
2	0	3	18
0	1	1	32
2	0	3	41
2	0	12	42
3	0	1	36
0	0	1	50
0	1	3	52
2	1	15	32
1	0	24	62
2	1	8	50
8	0	5	54
1	2	26	48
3	0	7	25
4	0	5	40
1	2	2	28
8	0	3	46
3	0	4	41
1	1	4	74
0	0	8	58
2	0	5	66
1	0	8	41
5	0	0	44
0	1	6	64
2	2	25	52

ANEXO 3

AMOSTRA DA PLANILHA DE ANÁLISE FATORIAL COM VALORES NORMALIZADOS E VARIÁVEIS REDUZIDAS

file	A	B	C	D	E	F	G	H	I	J	K	L	M				
	tokens	types	song	bundles	google	count	past	tense	present	tense	infinite	gerund	interj	qualifiers	possesives	nouns	prepositions
1	1940_fs_Apri	85	53	11,76	258,82	141,18	23,53	47,06	0,00	188,24	117,65	211,76	105,88				
2	1940_fs_Apri_played_a_fiddle.txt.token.tag																
3	1940_fs_Empty_saddies.txt.token.tag	107	56	0,00	93,46	28,04	130,84	65,42	0,00	261,68	37,38	205,61	140,19				
4	1940_fs_Every_night_about_this_time.txt.token.tag	124	50	32,26	201,61	0,00	96,77	56,45	2,42	241,94	137,10	225,81	129,03				
5	1940_fs_Hands_across_the_table.txt.token.tag	43	35	69,77	325,58	0,00	93,02	93,02	0,00	186,05	209,30	188,05	188,05				
6	1940_fs_Head_on_my_pillow.txt.token.tag	78	50	12,82	217,95	25,64	102,56	64,10	0,00	102,56	243,59	217,95	179,49				
7	1940_fs_ILL_Take_talulah.txt.token.tag	124	73	8,06	177,42	56,45	120,97	80,65	0,00	145,16	104,84	145,16	80,65				
8	1940_fs_ILL_never_let_a_day_pass_by.txt.token.tag	90	61	22,22	177,78	44,44	22,22	111,11	0,00	233,33	111,11	177,78	166,67				
9	1940_fs_ILL_never_smile_again.txt.token.tag	89	42	78,65	134,83	0,00	112,36	67,42	0,00	202,25	213,48	112,36	191,01				
10	1940_fs_ILL_sing_along_with_you.txt.token.tag	98	56	10,20	295,92	10,20	71,43	71,43	0,00	255,10	122,45	183,27	214,29				
11	1940_fs_I_could_make_you_care.txt.token.tag	78	46	38,46	256,41	102,56	25,64	166,87	0,00	115,38	294,87	128,21	76,92				
12	1940_fs_I_guess_ILL_have_to_dream_the_rest.txt.to	145	78	20,69	117,24	48,28	82,76	131,03	0,00	186,21	137,93	179,31	144,83				
13	1940_fs_I_havent_the_time_.txt.token.tag	92	62	54,35	206,52	43,48	54,35	65,22	0,00	282,61	86,96	239,13	141,30				
14	1940_fs_I_kissed_your_hand.txt.token.tag	57	37	0,00	157,89	17,54	87,72	70,18	0,00	140,35	210,53	238,25	105,26				
15	1940_fs_I_love_my_baby.txt.token.tag	121	60	8,26	173,55	0,00	123,97	74,38	0,00	223,14	239,67	181,82	33,06				
16	1940_fs_I_mamed_an_angel.txt.token.tag	67	45	14,93	164,18	119,40	59,70	29,85	0,00	283,58	164,18	164,18	104,48				
17	1940_fs_I_saw_your_face_in_a_cloud.txt.token.tag	131	64	15,27	359,78	53,44	83,97	22,90	0,00	190,84	229,01	213,74	122,14				
18	1940_fs_I_tried.txt.token.tag	81	56	37,04	185,19	160,49	61,73	49,38	0,00	160,49	234,57	180,49	98,77				
19	1940_fs_I_wanna_be_loved_by_you.txt.token.tag	189	37	63,49	211,64	63,49	10,58	164,02	0,00	126,98	211,64	232,80	105,82				
20	1940_fs_Id_know_anywhere.txt.token.tag	64	40	0,00	125,00	62,50	78,13	78,13	0,00	250,00	234,38	93,75	93,75				
21	1940_fs_Im_always_chasing_rainbows.txt.token.tag	120	74	16,67	259,33	33,33	116,67	66,67	0,00	258,33	108,33	200,00	91,67				
22	1940_fs_Im_in_love.txt.token.tag	317	78	138,80	258,68	78,96	56,78	9,46	0,32	176,66	160,88	211,36	164,04				
23	1940_fs_Im_in_the_mood_for_love.txt.token.tag	91	63	21,98	274,73	10,99	32,97	65,93	0,00	208,79	98,90	197,80	230,77				
24	1940_fs_In_a_little_Spanish_town.txt.token.tag	76	50	26,32	250,00	92,11	26,32	26,32	0,00	302,63	65,79	236,84	184,21				
25	1940_fs_Is_there_a_chance_for_me.txt.token.tag	170	50	5,88	394,12	11,76	94,12	64,71	0,00	264,71	135,29	217,65	117,65				
26	1940_fs_Ive_got_a_realless_spell.txt.token.tag	93	55	64,52	182,80	75,27	43,01	75,27	0,00	268,82	107,53	107,53	118,28				
27	1940_fs_Ive_got_my_eyes_on_you.txt.token.tag	51	35	19,61	196,08	78,43	39,22	98,04	0,00	176,47	196,08	137,25	176,47				
28	1940_fs_Ive_lost_my_heart_again.txt.token.tag	164	58	36,59	268,29	109,76	79,27	6,10	0,00	207,32	219,51	189,02	79,27				
29	1940_fs_Just_you_just_me.txt.token.tag	93	55	10,75	107,53	0,00	86,02	53,76	0,00	225,81	193,55	172,04	107,53				
30	1940_fs_Learn_to_croon.txt.token.tag	56	42	35,71	178,57	0,00	89,29	142,86	0,00	196,43	89,29	178,57	178,57				
31	1940_fs_The_sky_fell_down.txt.token.tag	71	55	14,08	309,86	140,95	42,25	14,08	0,00	211,27	169,01	197,18	126,76				



	possessives	nouns	prepositions	coord_conj	interrog	subord	modals	personal_pron	movement	emotion	musical_manifest	time	speech	people
1														
2	117,65	211,76	105,88	105,88	0,00	0,00	11,76	47,06	35,29	23,53	70,59	0,00	0,00	0,00
3	37,38	205,61	140,19	9,35	0,00	37,38	9,35	18,69	46,73	9,35	0,00	28,04	9,35	9,35
4	137,10	225,81	129,03	24,19	0,00	48,39	0,00	24,19	0,00	96,77	64,52	56,45	0,00	8,06
5	209,30	186,05	186,05	23,26	0,00	0,00	0,00	69,77	23,26	23,26	0,00	0,00	46,51	0,00
6	243,59	217,95	179,49	25,64	0,00	12,82	0,00	115,38	12,82	76,92	0,00	0,00	12,82	0,00
7	104,84	145,16	80,65	8,06	0,00	8,06	40,32	8,06	48,39	32,26	8,06	0,00	0,00	48,39
8	111,11	177,78	166,67	55,56	0,00	0,00	22,22	33,33	0,00	100,00	11,11	44,44	11,11	11,11
9	213,48	112,36	191,01	0,00	0,00	0,00	56,18	56,18	44,94	101,12	0,00	0,00	0,00	11,24
10	122,45	163,27	214,29	40,82	0,00	0,00	20,41	10,20	40,82	112,24	10,20	30,61	0,00	10,20
11	294,87	128,21	76,92	0,00	0,00	0,00	76,92	25,64	25,64	76,92	0,00	0,00	0,00	25,64
12	137,93	179,31	144,83	13,79	0,00	6,90	6,90	27,59	20,69	137,93	0,00	0,00	13,79	6,90
13	86,96	239,13	141,30	21,74	0,00	21,74	0,00	21,74	10,87	76,09	0,00	32,61	0,00	43,48
14	210,53	298,25	105,26	35,09	0,00	17,54	0,00	87,72	52,63	140,35	0,00	0,00	17,54	105,26
15	239,67	181,82	33,06	49,59	0,00	33,06	8,26	66,12	8,26	90,91	0,00	0,00	0,00	24,79
16	164,18	164,18	104,48	14,93	0,00	14,93	14,93	0,00	14,93	89,55	0,00	14,93	0,00	14,93
17	229,01	213,74	122,14	45,80	0,00	30,53	0,00	83,97	22,90	22,90	7,63	7,63	0,00	0,00
18	234,57	160,49	98,77	61,73	0,00	0,00	0,00	61,73	135,80	111,11	0,00	0,00	24,69	12,35
19	211,64	232,80	105,82	63,49	0,00	0,00	0,00	10,58	15,87	105,82	0,00	5,29	0,00	31,75
20	234,38	93,75	93,75	0,00	0,00	15,63	46,88	62,50	31,25	46,88	15,63	0,00	0,00	0,00
21	108,33	200,00	91,67	50,00	0,00	8,33	16,67	33,33	58,33	83,33	0,00	0,00	8,33	8,33
22	160,88	211,36	164,04	12,62	0,00	3,15	3,15	12,62	3,15	100,95	0,00	6,31	6,31	0,00
23	98,90	197,80	230,77	32,97	0,00	21,98	21,98	21,98	10,99	87,91	10,99	10,99	0,00	0,00
24	65,79	236,84	184,21	39,47	0,00	0,00	0,00	13,16	52,63	0,00	0,00	0,00	13,16	13,16
25	135,29	217,65	117,65	11,76	0,00	0,00	58,82	23,53	23,53	64,71	0,00	0,00	11,76	11,76
26	107,53	107,53	118,28	53,76	0,00	32,26	10,75	0,00	43,01	96,77	0,00	10,75	0,00	0,00
27	196,08	137,25	176,47	0,00	0,00	19,61	0,00	78,43	39,22	19,61	0,00	0,00	0,00	0,00
28	219,51	189,02	79,27	18,29	0,00	24,39	6,10	79,27	12,20	24,39	0,00	0,00	0,00	0,00
29	193,55	172,04	107,53	10,75	0,00	21,51	10,75	64,52	10,75	32,26	10,75	10,75	0,00	0,00
30	89,29	178,57	178,57	35,71	0,00	17,86	0,00	17,86	17,86	71,43	89,29	0,00	17,86	17,86
31	169,01	197,18	126,76	42,25	0,00	42,25	0,00	28,17	42,25	14,08	0,00	0,00	0,00	0,00

	object	social	interj
1			
2	11,76	0,00	0,00
3	9,35	9,35	0,00
4	16,13	104,84	2,42
5	0,00	23,28	0,00
6	25,64	12,82	0,00
7	8,06	16,13	0,00
8	0,00	44,44	0,00
9	0,00	78,65	0,00
10	0,00	10,20	0,00
11	0,00	38,46	0,00
12	6,90	13,79	0,00
13	0,00	32,61	0,00
14	17,54	70,18	0,00
15	0,00	41,32	0,00
16	14,93	14,93	0,00
17	7,63	0,00	0,00
18	0,00	49,38	0,00
19	0,00	52,91	0,00
20	0,00	0,00	0,00
21	0,00	25,00	0,00
22	0,00	0,00	0,32
23	0,00	0,00	0,00
24	0,00	13,16	0,00
25	0,00	11,76	0,00
26	10,75	32,28	0,00
27	0,00	19,61	0,00
28	0,00	0,00	0,00
29	0,00	86,02	0,00
30	0,00	17,86	0,00
31	0,00	28,17	0,00
32			

## ANEXO 4

LISTA DE PALAVRAS DE CLASSIFICAÇÃO SEMÂNTICA PARA A  
CATEGORIA ESTADO EMOCIONAL

===== EmotionalState =====

274 fear	12 uptight	5 tortured	3 worry	2 torment	1 furiously
188 shame	11 Sunshine	5 upset	2 amused	2 triumphant	1 heartstrings
165 lonely	11 Anger	4 bright	2 agony	2 unashamed	1 hectic
150 sunshine	11 emotional	4 grateful	2 afterglow	2 undertow	1 hysteria
117 thrill	11 glowing	4 gusto	2 Gall	2 zeal	1 indignation
106 sorry	10 excited	4 high	2 blissful	1 admiration	1 ingratitude
74 happiness	10 shadow	4 impatient	2 boot	1 alarm	1 intimidation
51 comfort	10 suffering	4 jubilation	2 brokenhearted	1 affect	1 irate
51 emotion	9 Thrill	4 overflow	2 cheerfulness	1 apprehension	1 joyfully
51 sadness	9 envy	4 pout	2 confident	1 apprehensive	1 loneliness
49 regret	9 laughing	4 spirit	2 disturbed	1 bad	1 mawkish
45 sorrow	8 craving	4 tense	2 ease	1 bang	1 misery
44 anger	8 lovesick	4 trouble	2 empathy	1 beaming	1 mournfully
44 sympathy	8 shiver	3 Bang	2 enmity	1 brood	1 nervy
39 bliss	7 Rue	3 anxiety	2 ennui	1 care	1 pained
36 ashamed	7 dread	3 animosity	2 esteem	1 casual	1 painfully
30 worried	7 lonesome	3 High	2 fevered	1 chafe	1 pinprick
28 wish	7 nervous	3 afraid	2 fit	1 chagrin	1 qualm
25 thankful	7 pleased	3 bitterness	2 gloat	1 chiller	1 raring
21 Wish	7 sadly	3 chill	2 glow	1 contrition	1 regretfully
21 jealousy	6 elation	3 contentment	2 grudge	1 creeps	1 repentance
20 numb	6 excitement	3 crazy	2 hopelessly	1 debonair	1 rush
20 surprise	6 frightened	3 gaily	2 irritation	1 dejected	1 sensation
19 mad	6 fuss	3 glee	2 itch	1 disgust	1 sentimental
18 excitation	6 passionate	3 gratitude	2 loathing	1 disgusted	1 serenity
18 panic	5 Cheer	3 grief	2 penitence	1 eagerly	1 shaken
15 cheer	5 Pain	3 jumpy	2 petulance	1 eagerness	1 shudder
15 gladness	5 Shiver	3 nonchalant	2 pissed	1 ecstatic	1 sinking
15 shy	5 anxious	3 shameless	2 premonition	1 electric	1 solace
14 Fear	5 boredom	3 spite	2 pride	1 enraged	1 sulk
14 remorse	5 caring	3 spoil	2 rue	1 enthusiasm	1 tightness
14 scared	5 compassion	3 state	2 rut	1 euphoria	1 unconcerned
13 fright	5 funk	3 tingle	2 shocked	1 euphoric	1 unsure
12 consolation	5 haunted	3 touched	2 tantrum	1 fire	1 wistful
12 pain	5 scare	3 wild	2 throes	1 foreboding	1 worship

## ANEXO 5

AMOSTRA DA PLANILHA DO EXCEL COM OS ESCORES PADRONIZADOS DOS COMPONENTES FATORIAIS PARA OS 30  
PRIMEIROS ARQUIVOS DO NOSSO CORPUS DE ESTUDO

1	file	Zposessi	Zpaast_te	Zpresent_te	Zinfinitive_gerur	Znouns	Zqualifier	Zprepositio	Zcoord_ci	Zinterrog
2	1940_fs_April_played_a_fiddle.txt.token	0,33	2,62	1,44	0,51	0,28	0,41	0,31	3,25	-
3	1940_fs_Empty_saddles.txt.token.tag	1,74	0,52	1,25	0,03	0,17	0,79	0,57	1,07	-
4	1940_fs_Every_night_about_this_time!	0,01	1,30	0,39	0,26	0,51	0,47	0,28	0,41	-
5	1940_fs_Hands_across_the_table.txt.t	1,27	1,30	0,30	0,71	0,15	0,45	1,74	0,45	-
6	1940_fs_Head_on_my_pillow.txt.token	1,87	0,59	0,54	0,06	0,38	1,82	1,57	0,34	-
7	1940_fs_ILL_Take_talulah.txt.token.ta	0,56	0,27	1,00	0,38	0,84	1,12	0,95	1,13	-
8	1940_fs_ILL_never_let_a_day_pass_by	0,45	0,07	1,47	1,19	0,29	0,32	1,25	1,00	-
9	1940_fs_ILL_never_smile_again.txt.tok	1,34	1,30	0,79	0,03	1,38	0,18	1,87	1,49	-
10	1940_fs_ILL_string_along_with_you.txt	0,25	1,02	0,24	0,13	0,53	0,68	2,46	0,34	-
11	1940_fs_I could_make_you_care.txt.to	2,76	1,55	1,39	2,66	1,12	1,61	1,05	1,49	-
12	1940_fs_I_guess_ILL_have_to_dream	0,02	0,04	0,04	1,71	0,27	0,45	0,69	0,87	-
13	1940_fs_I_havent_the_time_.txt.token	0,87	0,09	0,67	0,03	0,73	1,13	0,60	0,52	-
14	1940_fs_I_kissed_your_hand.txt.token	1,29	0,81	0,17	0,10	1,72	1,20	0,32	0,08	-
15	1940_fs_I_love_my_baby.txt.token.tag	1,80	1,30	1,08	0,21	0,22	0,16	2,17	0,73	-
16	1940_fs_I_married_an_angel.txt.token.	0,48	2,01	0,54	0,97	0,52	1,15	0,34	0,82	-
17	1940_fs_I_saw_your_face_in_a_cloud.t	1,61	0,18	0,07	1,15	0,31	0,37	0,11	0,56	-
18	1940_fs_I_tried.txt.token.tag	1,71	3,15	0,48	0,45	0,58	0,87	0,49	1,27	-
19	1940_fs_I_wanna_be_loved_by_you.tx	1,31	0,46	1,77	2,59	0,63	1,42	0,31	1,35	-
20	1940_fs_Id_know_anywhere.txt.token.	1,71	0,44	0,07	0,31	1,69	0,60	0,62	1,49	-
21	1940_fs_Im_always_chasing_rainbows	0,50	0,37	0,89	0,01	0,08	0,73	0,67	0,75	-
22	1940_fs_Im_in_love.txt.token.tag	0,42	0,89	0,61	1,51	0,27	0,60	1,18	0,93	-
23	1940_fs_Im_in_the_mood_for_love.txt.	0,66	0,99	1,21	0,01	0,04	0,08	2,88	0,01	-
24	1940_fs_In_a_little_Spanish_town.txt.t	1,24	1,26	1,37	1,06	0,69	1,46	1,69	0,28	-
25	1940_fs_Is_there_a_chance_for_me.tx!	0,03	0,97	0,33	0,04	0,37	0,84	0,01	0,97	-
26	1940_fs_Ive_got_a_restless_spell.txt.t	0,51	0,79	0,95	0,24	1,46	0,91	0,01	0,92	-
27	1940_fs_Ive_got_my_eyes_on_you.txt.	1,04	0,88	1,05	0,84	0,97	0,61	1,50	1,49	-
28	1940_fs_Ive_lost_my_heart_again.txt.t	1,45	1,75	0,04	1,60	0,10	0,10	0,99	0,67	-
29	1940_fs_Just_you_just_me.txt.token.ta	0,99	1,30	0,13	0,33	0,39	0,20	0,27	1,01	-
30	1940_fs_Learn_to_croon.txt.token.tag	0,83	1,30	0,21	2,03	0,28	0,28	1,55	0,11	-



1	Zsubord	Zmodals	Zpersonal_p	Zinterj	Zmoveme	Zemotion	Zmusical	Ztime	Zspeech	Zpeople	Zobject	Zsocial	Zsongbun	Zgoogle_cou
2	0,75	0,06	0,74	0,39	0,11	0,88	2,62	0,68	0,62	0,61	0,15	1,11	0,46	0,38
3	1,83	0,22	0,47	0,39	0,47	1,23	0,37	0,58	0,04	0,29	0,01	0,73	0,86	1,59
4	2,59	0,82	0,23	1,10	1,00	0,91	2,37	1,86	0,62	0,34	0,40	3,08	0,24	0,30
5	0,75	0,82	1,71	0,39	0,27	0,89	0,37	0,68	2,68	0,61	0,53	0,18	1,51	1,18
6	0,14	0,82	3,65	0,39	0,59	0,42	0,37	0,68	0,29	0,61	0,96	0,60	0,42	0,11
7	0,19	1,76	0,92	0,39	0,52	0,67	0,02	0,68	0,62	1,04	0,06	0,46	0,58	0,59
8	0,75	0,60	0,16	0,39	1,00	0,99	0,11	1,32	0,17	0,23	0,53	0,67	0,10	0,59
9	0,75	2,77	1,13	0,39	0,41	1,02	0,37	0,68	0,62	0,23	0,53	2,04	1,81	1,10
10	0,75	0,49	0,83	0,39	0,28	1,29	0,07	0,70	0,62	0,26	0,53	0,70	0,51	0,82
11	0,75	4,10	0,17	0,39	0,19	0,42	0,37	0,68	0,62	0,26	0,53	0,43	0,45	0,35
12	0,27	0,38	0,09	0,39	0,35	1,92	0,37	0,68	0,36	0,38	0,13	0,56	0,16	1,31
13	0,75	0,82	0,34	0,39	0,66	0,40	0,37	0,79	0,62	0,87	0,53	0,20	0,98	0,24
14	0,46	0,82	2,47	0,39	0,65	1,97	0,37	0,68	0,62	2,98	0,49	1,70	0,86	0,83
15	1,53	0,29	1,55	0,39	0,74	0,77	0,37	0,68	0,62	0,23	0,53	0,54	0,58	0,64
16	0,28	0,14	1,26	0,39	0,53	0,73	0,37	0,01	0,62	0,10	0,33	0,51	0,35	0,75
17	1,36	0,82	2,31	0,39	0,28	0,90	0,04	0,33	0,62	0,61	0,09	1,11	0,34	1,57
18	0,75	0,82	1,36	0,39	3,26	1,26	0,37	0,68	1,13	0,19	0,53	0,87	0,40	0,50
19	0,75	0,82	0,81	0,39	0,50	1,13	0,37	0,44	0,62	0,47	0,53	1,01	1,29	0,18
20	0,33	2,18	1,40	0,39	0,02	0,31	0,30	0,68	0,62	0,61	0,53	1,11	0,86	1,22
21	0,17	0,25	0,16	0,39	0,83	0,58	0,37	0,68	0,68	0,33	0,53	0,11	0,29	0,37
22	0,53	0,61	0,73	0,20	0,90	1,01	0,37	0,39	0,17	0,61	0,53	1,11	3,85	0,38
23	0,77	0,59	0,33	0,39	0,65	0,69	0,10	0,18	0,62	0,61	0,53	1,11	0,11	0,57
24	0,75	0,82	0,70	0,39	0,65	1,46	0,37	0,68	0,31	0,16	0,53	0,58	0,03	0,28
25	0,75	2,94	0,26	0,39	0,26	0,13	0,37	0,68	0,21	0,21	0,53	0,64	0,66	2,00
26	1,48	0,13	1,26	0,39	0,35	0,91	0,37	0,19	0,62	0,61	0,09	0,18	1,33	0,53
27	0,61	0,82	2,07	0,39	0,23	0,98	0,37	0,68	0,62	0,61	0,53	0,32	0,19	0,37
28	0,93	0,43	2,11	0,39	0,61	0,86	0,37	0,68	0,62	0,61	0,53	1,11	0,38	0,49
29	0,74	0,13	1,48	0,39	0,66	0,67	0,09	0,19	0,62	0,61	0,53	2,33	0,49	1,43
30	0,48	0,82	0,50	0,39	0,44	0,29	3,42	0,68	0,65	0,00	0,53	0,39	0,35	0,58



## ANEXO 6

## CATEGORIZAÇÃO DOS 32 ARTISTAS DO NOSSO CORPUS DE ESTUDO POR ESTILO MUSICAL

ARTISTA	complemento características, influências e estilos predominantes (Billboard)	ESTILO / FONTE				NOSSO
		Starr; Waterman, 2007 e Charlton, 2009)	allmusic e musicimprint	Rolling Stone Magazine		
<b>AEROSMITH</b>	Banda americana de “ <i>hard rock</i> ”; Seu estilo tem raízes do blues, incorporando elementos do pop, R&B, heavy metal ( <b>Hard rock</b> , blues-rock, heavy metal)	<b>hard rock</b>	Pop/Rock	<b>Hard Rock</b> ; Heavy Metal; Album Rock; Pop/Rock; Arena Rock; Pop-Metal	<b>HARD ROCK</b>	
<b>BEACH BOYS</b>	Americanos, precursores do estilo <b>Surf rock</b> , sunshine pop, psychedelic pop, psychedelic rock, baroque pop, art rock	rock	Pop/Rock	AM Pop; Early Pop/Rock; <b>Surf</b> ; Pop/Rock; Sunshine Pop; Rock & Roll; Psychedelic Pop; Psychedelic	<b>SURF ROCK</b>	
<b>BEATLES</b>	Britânicos, com estilo inicial enraizado no rock ‘n’ roll dos anos 50, seguido de diversas influências como Tin Pan Alley, pop, rock psicodélico. Visual colorido, foram grande influência social e cultural nos anos 60. <b>Rock/pop</b>	rock	<b>Pop/Rock</b>	Early <b>Pop/Rock</b> ; Rock & Roll; British Invasion; Psychedelic; Merseybeat; AM Pop; <b>Pop/Rock</b> ; British Psychedelia; Folk- Rock	<b>POP ROCK</b>	
<b>BON JOVI</b>	banda de <b>rock</b> americana	<b>hard rock</b>	Pop/Rock	<b>Hard Rock</b> ; Pop/Rock; Hair Metal; Pop-Metal; Arena Rock; Heavy Metal; Album Rock	<b>HARD ROCK</b>	
<b>CREED</b>	banda de <b>rock</b> Americana de Tallahassee, FL, com influência <b>grunge</b> .	<b>post-grunge</b>	Pop/Rock	<b>Post-Grunge</b> ; Alternative Pop/Rock	<b>POST GRUNGE</b>	
<b>DEF LEPPARD</b>	Britânicos, com estilos que envolvem do <b>Hard rock</b> ; heavy metal e Glam Metal	rock (English rock)	Pop/Rock	British Metal; <b>Hard Rock</b> ; Pop/Rock; Heavy Metal; Hair Metal; Pop- Metal; New Wave of British Heavy Metal	<b>HARD ROCK</b>	

ARTISTA	complemento características, influências e estilos predominantes (Billboard)	ESTILO / FONTE				NOSSO
		Starr; Waterman, 2007 e Charlton, 2009)	allmusic e musicimprint	Rolling Stone Magazine	POP	
<b>ELTON JOHN</b>	Britânico, cujas músicas exerceram forte impacto na música popular, especialmente para o rock tocado no piano. Rock, <b>pop</b> , glam rock	<b>pop</b>	Pop/Rock	Rock & Roll; Singer/Songwriter; Adult Contemporary; Soft Rock; <b>Pop</b> ; Pop/Rock; Album Rock; AM Pop	<b>POP</b>	
<b>ELVIS PRESLEY</b>	Americano, iniciou a carreira com o estilo rockabilly, com influências de R&B e uma forte batida, misturando música negra e branca. <b>rock'n'roll</b> , gospel, blues, country, ballads and pop, Rockabilly, Blues, Gospel, R&B	<b>rock'n'roll</b>	Pop/Rock	<b>Rock &amp; Roll</b> ; Rockabilly; Early Pop/Rock; Pop/Rock; AM Pop	<b>ROCK'N' ROLL</b>	
<b>FOO FIGHTERS</b>	Banda americana de <b>rock</b> , fundada no período pós grunge	rock	Pop/Rock	<b>Post-Grunge</b> ; Alternative Pop; Grunge	<b>POST GRUNGE</b>	
<b>FRANK SINATRA</b>	Americano, iniciou a carreira a era do swing, atingindo sucesso como artista solo em meados dos anos 40. <b>Traditional pop</b> , jazz, big band, <b>vocal</b>	<b>pop</b> (traditional)	<b>VOCAL/POP</b>	Ballads; Swing; Standards; Traditional <b>Pop</b> ; <b>Vocal Jazz</b> ; Soft Rock; Show Tunes; AM Pop	<b>POP (VOCAL)</b>	
<b>GREENDAY</b>	Americanos <b>Punk rock</b> , pop punk, alternative rock	<b>punk rock</b>	Pop/Rock	Alternative Pop/ Rock; Punk Revival; Post-Grunge; <b>Punk Rock</b>	<b>PUNK ROCK</b>	
<b>IRON MAIDEN</b>	Banda de <b>heavy metal</b> inglesa, formada em 1975.	<b>heavy metal</b>	Pop/Rock	<b>Heavy Metal</b> ; British Metal; New Wave of British Heavy Metal	<b>HEAVY METAL</b>	
<b>JOHNNY CASH</b>	Compositor e cantor americano, um dos artistas mais influentes do século 20. <b>Country</b> , rock'n'roll, folk, gospel, blues, rockabilly	<b>country</b>	<b>COUNTRY</b>	Traditional <b>Country</b> ; Cowboy; Country-Pop; Rockabilly; Country Gospel; Rock & Roll	<b>COUNTRY</b>	
<b>KILLERS</b>	Banda americana, com forte influência do pop dos anos 80. <b>Alternative rock</b> , post-punk revival	<b>alternative rock</b>	Pop/Rock	<b>Alternative Pop/Rock</b> ; New Wave/ Post-Punk Revival	<b>ALTERNATIVE ROCK</b>	

ARTISTA	complemento características, influências e estilos predominantes (Billboard)	ESTILO / FONTE				NOSSO
		Starr; Waterman, 2007 e Charlton, 2009)	allmusic e musicimprint	Rolling Stone Magazine		
<b>LED ZEPPELIN</b>	Banda inglesa formada em 1968 que apresentou uma nova interpretação para os energos folk e blues, com influências de rockabilly, reggae, funk, música do oriente. Hard rock, heavy metal, folk rock, <b>rock</b>	blues rock	Pop/Rock	<b>Rock</b> ; Blues-Rock; Hard Rock; Heavy Metal; Arena Rock; British Metal; British Blues; Album Rock	<b>ROCK</b>	
<b>LENNY KRAVITZ</b>	Cantor e compositor americano, multi-instrumentista, com um estilo “retro” que engloba elementos do <b>rock</b> , soul, funk, reggae, hard rock, música psicodélica, hard rock, pop rock, R&B e blues-rock	blues rock	Pop/Rock	<b>Rock</b> ; Neo-Psychedelia; American Trad Rock; Album Rock	<b>ROCK</b>	
<b>MADONNA</b>	Cantora americana, rainha do pop <b>Pop</b> , rock, dance, música eletrônica	<b>POP</b>	<b>Pop/Rock</b>	<b>Pop</b> ; Dance-Pop; Adult Contemporary; Pop/Rock; Club/Dance	<b>POP</b>	
<b>METALLICA</b>	banda de heavy metal americana <b>Heavy metal</b> , thrash metal, hard rock, speed metal	<b>heavy Metal</b>	Pop/Rock	Thrash; <b>Heavy Metal</b> ; Speed Metal; Hard Rock	<b>HEAVY METAL</b>	
<b>MICHAEL JACKSON</b>	Cantor americano, também conhecido como o “rei do pop”, exerceu forte influência na música comercial americana. R&B, <b>pop</b> , rock, dance	<b>pop</b>	R & B	<b>Pop</b> ; Dance-Pop; Urban; Motown; Pop/Rock; Club/Dance; New Jack Swing; AM Pop	<b>POP</b>	
<b>MCR (My Chemical Romance)</b>	Banda de rock americana, formada em 2001. Emo; Alternative rock; Post-hardcore; <b>Pop punk</b>	alternative rock; emo	<b>Pop/Rock</b>	Punk Revival; Post-Hardcore; <b>Punk-Pop</b> ; Alternative Pop/ Rock; Emo	<b>PUNK POP</b>	
<b>NICKELBACK</b>	banda de rock canadense, formada em 1995. Hard rock; <b>Post-grunge</b> ; Metal alternativo; rock alternativo; pop rock	alternative rock	Pop/Rock	Alternative Metal; <b>Post-Grunge</b>	<b>POST GRUNGE</b>	
<b>NIRVANA</b>	banda re rock americana dos anos 90, precursores do grunge. Alternative rock, <b>grunge</b>	<b>grunge</b>	Pop/Rock	Alternative Pop/ Rock; <b>Grunge</b>	<b>GRUNGE</b>	

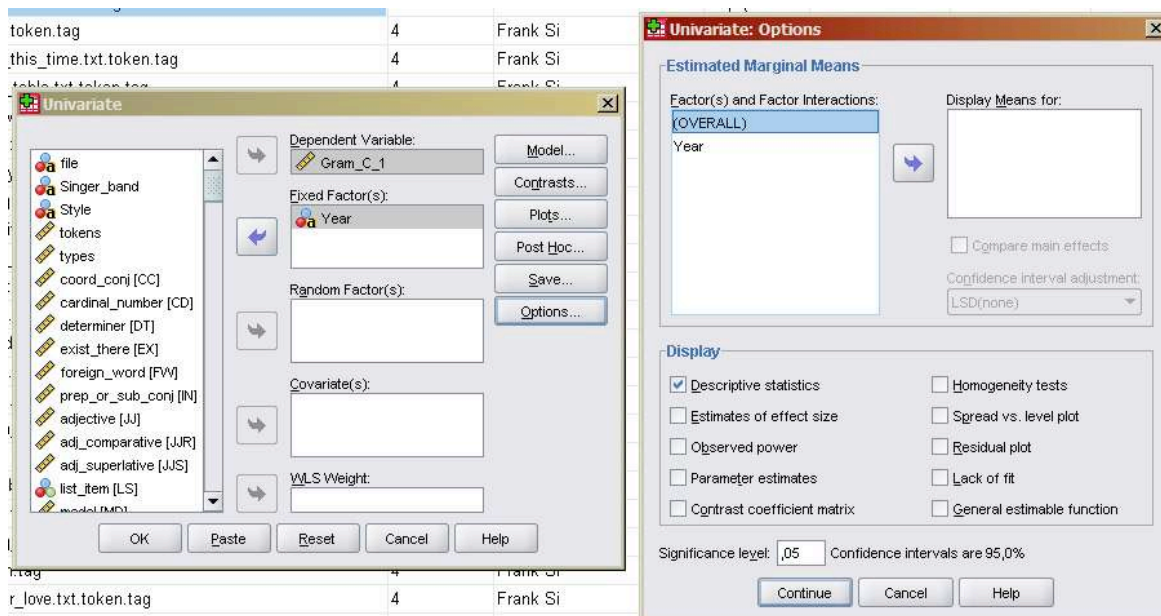
ARTISTA	complemento características, influências e estilos predominantes (Billboard)	ESTILO / FONTE			NOSSO
		Starr; Waterman, 2007 e Charlton, 2009)	allmusic e musicimprint	Rolling Stone Magazine	
<b>PARAMORE</b>	banda de rock americana, formada no Tennessee em 2004. alternative rock, Emo, <b>Pop punk</b>	alternative rock	Pop/Rock	Emo-Pop; Alternative Pop/ Rock; <b>Punk-Pop</b> ; Pop	<b>PUNK POP</b>
<b>PEARL JAM</b>	Banda de rock americana, formada em Seattle, em 1990. Alternative rock, <b>grunge</b> , hard rock	<b>grunge</b>	Pop/Rock	Hard Rock; Alternative Pop/ Rock; <b>Grunge</b>	<b>GRUNGE</b>
<b>PINK FLOYD</b>	banda inglesa, do final dos anos 60, reconhecidos por sua experimentação sonora e letras filosóficas. <b>Progressive rock</b> , psychedelic rock	<b>progressive rock</b>	Pop/Rock	<b>Prog-Rock/ Art Rock</b> ; British Invasion; Psychedelic; Album Rock; British Rock; Psychedelia; Hard Rock	<b>PROGRESSIVE ROCK</b>
<b>QUEEN</b>	banda inglesa, formada em 1970. <b>Rock</b> , hard rock	<b>rock</b>	Pop/ <b>Rock</b>	Hard Rock; Prog-Rock/ Art Rock; Glam Rock; Pop/ <b>Rock</b> ; Album Rock; Dance-Rock; Arena Rock; British Metal	<b>ROCK</b>
<b>QUEENSRYCHE</b>	Banda americana, formada em 1981. Heavy metal, <b>progressive metal</b>	heavy metal	Pop/Rock	<b>Prog-Rock/ Art Rock</b> ; Heavy Metal; <b>Progressive Metal</b> ; Hard Rock; Album Rock; Arena Rock	<b>PROGRESSIVE ROCK</b>
<b>RAY CHARLES</b>	Cantor e compositor americano que incorporou o estilo soul ao country e ao rock 'n' roll. Rhythm and blues, <b>soul</b> , rock 'n' roll, blues, jazz, <b>country</b> , pop	<b>country soul</b>	R&B	Soul; Early R&B; Blues; Piano Blues; <b>Country-Soul</b> ; Pop-Soul; Early Pop/Rock; Urban Blues; Jazz Blues; AM Pop	<b>COUNTRY SOUL</b>
<b>RED HOT</b>	Banda de rock americana, californiana, formada em 1983, com estilo que mistura elementos do funk, do punk <b>alternativo</b> , punk rock e música psicodélica.	<b>alternative rock</b>	Pop/Rock	Alternative Pop/ Rock; Funk Metal; College Rock; Dance-Rock; <b>Alternative Rock</b> ; Rap-Rock	<b>ALTERNATIVE ROCK</b>

ARTISTA	complemento características, influências e estilos predominantes (Billboard)	ESTILO / FONTE			
		Starr; Waterman, 2007 e Charlton, 2009)	allmusic e musicimprint	Rolling Stone Magazine	NOSSO
<b>ROLLING STONES</b>	banda inglesa, formada em 1962 <b>Rock</b> , rock,n, roll,R&B, blues rock, blues	rock	Pop/Rock	Rock & Roll; Blues-Rock; Hard Rock; British Invasion; <b>Rock</b> ; British Blues; Album Rock; Psychedelic; British Psychedelia; Dance-Rock; Early Pop/Rock; AM Pop	<b>ROCK</b>
<b>SIMPLE PLAN</b>	banda canadense, formada no ano 2000 <b>pop punk</b>	<b>pop punk</b>	Pop/Rock	<b>Punk-Pop</b> Punk Revival	<b>PUNK POP</b>
<b>U2</b>	banda britânica, formada na Irlanda nos anos 80. <b>pop, rock</b> , alternative rock, post-punk	rock	<b>Pop/Rock</b>	Alternative Pop/ Rock <b>Pop/Rock</b> Post-Punk College Rock Dance-Rock	<b>POP ROCK</b>

## ANEXO 7

IMAGEM DA ANÁLISE DESCRITIVA FINAL NO PROGRAMA SPSS PARA  
CADA COMPONENTE FATORIAL SOB A PERSPECTIVA “YEAR”

A análise descritiva final computa os pesos de cada componente de cada fator (Gram\_C1; Gram\_C2; Gram\_C3; Sem\_C1; Sem\_C2; Sem\_C3; LexGram\_C1; LexGram\_C2; LexGram\_C3) para cada um dos textos do corpus, a partir da perspectiva selecionada: ano (*year*); artista (*singer\_band*); ou estilo (*style*). O quadro abaixo ilustra este procedimento no programa SPSS para o fator Gram\_C1 sob a perspectiva ano. O SPSS retorna tabelas com os textos que possuem as características mais e menos representantes para cada fator, as quais podem ser salvas como planilhas do Excel, que possibilitaram a análise exposta no item 5.2.



## ANEXO 8

A seguir estão expostos quadros com as letras de música mais frequentes na dimensão gramatical 1, considerado-se as perspectivas “ano”, “tempo” e “estilo”. Dada a dimensão de nossos dados apresentamos apenas as cinco letras de música mais representativas e as cinco letras de música menos representativas.

file	singer	gram_c1
2002_sp_grow_up.txt.token.tag	Simple Plan	4,41
2008_sp_crash_and_burn.txt.token.tag	Simple Plan	3,87
2008_sp_save_you.txt.token.tag	Simple Plan	3,73
2004_sp_perfect_world.txt.token.tag	Simple Plan	3,53
2008_sp_happy_together.txt.token.tag	Simple Plan	3,28
1975_aero_no_more.txt.token.tag	Aerosmith	-4,17
1982_aero_jig_is_up.txt.token.tag	Aerosmith	-4,34
1975_aero_toys_in_the_attic.txt.token.tag	Aerosmith	-4,79
1977_aero_bright_light_fright.txt.token.tag	Aerosmith	-4,82
1985_aero_shame_on_you.txt.token.tag	Aerosmith	-5,68

file	year	gram_c1
2005_alter_you_will_soar.txt.token.tag	0	7,62
2005_mad_fighting_spirit.txt.token.tag	0	7,21
2003_im_judgement_day.txt.token.tag	0	7,2
2002_nic_I_will_always_follow_you.txt.token.tag	0	6,87
2009_para_my_number_one.txt.token.tag	0	6,68
1985_aero_shame_on_you.txt.token.tag	8	-5,68
1984_bon_roulette.txt.token.tag	8	-5,89
1986_bon_raise_your_hands.txt.token.tag	8	-5,93
1985_bon_always_run_to_you.txt.token.tag	8	-6,31
1987_pfw_radio_waves.txt.token.tag	8	-6,57

file	style	gram_c1
2009_para_my_number_one.txt.token.tag	PUNK POP	6,68
2009_para_swim_in_silence.txt.token.tag	PUNK POP	5,91
2008_sp_I_can_wait_forever.txt.token.tag	PUNK POP	5,71
2004_sp_how_could_this_happen_to_me.txt.toke	PUNK POP	5,36
2002_sp_my_alien.txt.token.tag	PUNK POP	5,28
1964_bb_pom_pom_play_girl.txt.token.tag	ROCK (surf)	-5,61
1964_bb_santas_got_an_airplane.txt.token.tag	ROCK (surf)	-5,03
1964_bb_little_saint_nick.txt.token.tag	ROCK (surf)	-5
1963_bb_lonely_sea.txt.token.tag	ROCK (surf)	-4,61
1963_bb_Spirit_of_America.txt.token.tag	ROCK (surf)	-4,51

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)



[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)