



Barbara Cristina Marques Pereira Ramos

**Metodologias para caracterização e detecção da descrição
de emoção na língua portuguesa**

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de Doutora em Letras/Estudos Estudos da Linguagem pelo Programa de Pós-graduação em Estudos da Linguagem da PUC-Rio.

Orientadora: Maria Cláudia de Freitas

Coorientadora: Diana Maria de Sousa Marques Pinto dos Santos

Rio de Janeiro
Abril 2023



Barbara Cristina Marques Pereira Ramos

**Metodologias para caracterização e detecção
da descrição de emoção na língua portuguesa**

Tese apresentada como requisito parcial para obtenção do grau de Doutora pelo Programa de Pós-graduação em Estudos da Linguagem da PUC-Rio. Aprovada pela Comissão Examinadora abaixo:

Maria Cláudia de Freitas

Orientadora

Departamento de Letras – PUC-Rio

Diana Maria de Sousa Marques Pinto dos Santos

Coorientadora

University of Oslo

Adriana Nogueira Accioly Nóbrega

Departamento de Letras – PUC-Rio

Oto Araújo Vale

UFSCAR

Emanoel Cesar Pires de Assis

UEMA

Belinda Mary Harper Sousa Maia

UP

Rio de Janeiro, 19 de abril de 2023.

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, da autora e do orientador.

Barbara Cristina Marques Pereira Ramos

Graduada em Letras pela Universidade do Estado do Rio de Janeiro (2014), mestra em Letras pela UERJ (2016) e doutora em Estudos da Linguagem pelo PPGEL/PUC-Rio (2023) com doutorado sanduíche na Universidade de Oslo (2019-2020). É professora de língua inglesa atuante no Ensino Fundamental Anos Iniciais como professora regente em escola bilíngue.

Ficha Catalográfica

Ramos, Barbara Cristina Marques Pereira

Metodologias para caracterização e detecção da descrição de emoção na língua portuguesa / Barbara Cristina Marques Pereira Ramos ; orientadora: Maria Cláudia de Freitas ; coorientadora: Diana Maria de Sousa Marques Pinto dos Santos. – 2023.

138 f. : il. color. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Letras, 2023.

Inclui bibliografia

1. Letras – Teses. 2. Descrição do português. 3. Descrição de emoção. 4. Estudos linguísticos com corpus. 5. Processamento de linguagem natural. 6. Anotação linguística. I. Freitas, Maria Cláudia de. II. Santos, Diana Maria de Sousa Marques Pinto dos. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Letras. IV. Título.

CDD: 400

Agradecimentos

A Deus, por sua grandiosidade e abrigo, e a meus queridos e amados Mestres, por sempre me acompanharem.

A meus pais, Robson e Márcia, pela educação, confiança e apoio incondicionais. A meus irmãos, Brian e Renata, pelo companheirismo e exemplo de dedicação e comprometimento. A minha avó Adélia, pelo amor que sempre me deu. A meus padrinhos, Solange e Paulo, e minha prima Karla, por todo o carinho e incentivo de sempre. A meus sogros, Célia e Vicente, pelo apoio e acolhimento.

A meu marido, Fábio, pelo amor, respeito, parceria e apoio diários. A meu filho, Antônio, por me mostrar um novo sentido para a vida.

A meus amigos: pelos conselhos e pelas palavras de afeto e motivação.

Às minhas queridas orientadoras, que acreditaram em mim mais que eu mesma e me deram apoio e encorajamento para seguir em frente.

Aos professores que compuseram minha banca examinadora, por terem aceitado o convite para participar da minha defesa.

À PUC-Rio e à CAPES, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Ramos, Barbara Cristina Marques Pereira; Freitas, Maria Cláudia de (Orientadora); Santos, Diana (Coorientadora). **Metodologias para caracterização e detecção da descrição de emoção na língua portuguesa**. Rio de Janeiro, 2023. 138p. Tese de Doutorado – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

O interesse desta tese recai sobre compreender como os falantes de língua portuguesa a utilizam para materializar a menção de emoção através de um trabalho, sobretudo, linguístico. O objetivo geral da pesquisa é criar recursos para aprimorar a anotação do campo semântico das emoções na língua portuguesa a partir do projeto AC/DC, projeto que reúne e disponibiliza publicamente corpora anotados e recursos para pesquisas na língua portuguesa, e do Emocionário, projeto de anotação semântica e léxico de emoções. Inicialmente, a pesquisa dá um panorama dos estudos de emoção; se alinha às perspectivas que refutam a universalidade das emoções e abordagens que postulam emoções básicas; e contrapõe seu interesse por menção de emoção à já consolidada área de Análise de Sentimento, contrastando cinco léxicos de sentimento e/ou polaridades em língua portuguesa e o Emocionário. A partir de uma ampla varredura nos corpora do AC/DC, três principais caminhos foram percorridos para investigar palavras de emoção: (i) uma análise dos vinte e quatro grupos de emoção que já existiam no léxico do Emocionário a fim de delinear características e desafios no estudo de emoção na língua portuguesa; (ii) a revisão completa um terço dos grupos do léxico do Emocionário; e (iii) buscas pelo padrão léxico-sintático “sentimento de N” e por expressões anotadas pelo projeto Esqueleto usadas para descrever emoção. A análise dos corpora à luz dos lemas previamente pertencentes aos grupos do léxico do Emocionário evidenciou, dentre outras características, a relevância de expressões lexicalizadas para a análise da descrição de emoção, dos tipos de argumentos de verbos e afixos que podem causar variação de sentido, e

de variações de tempo e modo verbal que acarretam mudança de significado. Dentre os desafios estão palavras e expressões polissêmicas e a dificuldade na detecção de diferentes sentidos em palavras que compartilham da mesma classe gramatical, tendo como base somente informações morfossintáticas. Esta análise possibilitou a estruturação e documentação de uma metodologia de revisão que pode vir a ser aplicada nos demais grupos futuramente. As principais contribuições desta tese são decorrentes das análises e explorações em corpora: a limpeza de lemas com sentidos não-emocionais dos grupos do léxico do Emocionário; a criação dos grupos de emoção “Ausência” e “Outra”, enriquecendo o léxico; a detecção de mais de novecentos lemas e expressões provenientes das buscas pelo padrão “sentimento de N” e das conexões estabelecidas entre os campos semânticos de emoção e do corpo humano; além de descobertas de campos lexicais pouco mencionados na literatura sobre emoção, como “coletividade”, “estranhamento”, “espiritualidade”, “parentesco” e “atos automotivados”, que auxiliaram na investigação de como os falantes do português cristalizam emoções na língua.

Palavras-chave

Descrição do português; Descrição de emoção; Estudos linguísticos com corpus; Processamento de linguagem natural; Anotação linguística; Léxicos.

Abstract

Ramos, Barbara Cristina Marques Pereira; Freitas, Maria Cláudia de (Advisor); Santos, Diana (Coadvisor). **Methodologies for characterizing and detecting emotional description in the Portuguese language**. Rio de Janeiro, 2023. 138p. Tese de Doutorado – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

The interest of this thesis lies in understanding how Portuguese speakers use it to materialize the mention of emotion through a linguistic perspective. The general objective of the research is to create resources to improve the annotation of the semantic field of emotions in the Portuguese language based on the AC/DC project, which gathers and makes publicly available annotated corpora and tools for linguistic research on Portuguese language. and Emocionário, which is both a semantic annotation project and lexicon of emotions. Initially, the research gives an overview of emotion studies; aligning itself with perspectives that refute the universality of emotions and approaches that postulate basic emotions; and contrasts the interest in emotion description to the already consolidated area of Sentiment Analysis, comparing five lexicons of emotion and/or polarities in Portuguese to Emocionário. From a broad sweep of the AC/DC corpora, three main paths were taken towards investigating emotion words: (i) an analysis of the twenty-four emotion groups previously composing the Emocionário lexicon in order to delineate characteristics and challenges in the study of emotion description in the Portuguese language; (ii) a thorough revision of one-third of the Emocionário lexicon groups; and (iii) searches for the lexical-syntactic pattern “sentimento de N” and for expressions annotated by the Esqueleto project used to describe emotion. The corpora analysis in the light of the lemmas previously belonging to the Emocionário lexicon groups showed, amongst other characteristics, the relevance of lexicalized expressions for the analysis of the emotion description, the types of arguments of verbs and affixes that can cause

variation in meaning, and variations in tense and verbal mode that lead to a change in meaning. Amongst the challenges are polysemous words and expressions and the difficulty in detecting different meanings in words that share the same grammatical class, based only on morphosyntactic information. This analysis enabled the structuring and documentation of a revision methodology that may be applied in other groups in the future. The main contributions of this thesis derive from the analyzes and explorations in corpora: the exclusion of lemmas with non-emotional meanings from the Emocionário lexicon groups; the creation of emotion groups “Ausência” and “Outra”, enriching the lexicon; the detection of more than nine hundred lemmas and expressions from the searches for the “sentimento de N” pattern and the connections established between the semantic fields of emotion and the human body; in addition to discoveries of lexical fields rarely mentioned in the literature on emotion, such as “coletividade”, “estranhamento”, “espiritualidade”, “parentesco” e “atos automotivados”, which helped in the investigation of how Portuguese speakers crystallize emotions in language.

Keywords

Portuguese description; Emotion analysis; Corpora analysis; Natural Language Processing; Linguistic annotation; Lexicons.

Sumário

1	Introdução.....	1
1.1	Motivação: estudo de emoções na língua	1
1.2	Contexto	6
1.3	Objetivos	6
2	Áreas de pesquisa: enquadramento teórico metodológico	9
2.1	Estudos linguísticos com corpus	9
2.2	AC/DC, Linguateca e Gramateca	11
2.3	O Emocionário.....	13
2.4	Humanidades Digitais e leitura distante	21
2.5	Aplicação prática da leitura distante ao estudo de emoções em corpus	22
3	O Emocionário e Léxicos de sentimento em português: Análise de menção de emoção é Análise de Sentimento?	25
3.1	Análise de Sentimento.....	25
3.2	Léxicos de sentimento em língua portuguesa	28
3.2.1	WordNetAffectBR.....	28
3.2.2	OPLexicon.....	30
3.2.3	LIWC / <i>Brazilian Portuguese LIWC 2007 Dictionary</i>	31
3.2.4	SentiLex-PT	33
3.2.5	Multilingualsentiment.....	34
3.3	Por que não é possível comparar os léxicos ao Emocionário	37
4	Características e desafios da análise de emoção.....	39
4.1	Expressões idiomáticas e expressões de várias palavras.....	39
4.2	Polissemia	41
4.2.1	Palavras polissêmicas cujos sentidos variam entre palavra de emoção ou não.....	42
4.2.2	Polissemia de sentidos de emoção em uma mesma palavra:.....	43
4.3	Classes gramaticais e a variação de sentido emocional	44
4.4	Argumentos do verbo	47
4.5	Palavras de emoção ocorrendo em um contexto de negação.....	49

4.6	Morfologia: prefixos e sufixos em palavras de emoção	50
4.7	Mudanças de tempo e modo verbal	50
4.8	O peso emocional das palavras de emoção.....	52
5	Ampliação e validação do Emocionário	54
5.1	“Sentimento de N”	56
5.1.1	Tratamento da cauda longa do “sentimento de N”	60
5.2	O mapa emocional do corpo.....	66
6	Revisão e reorganização do léxico: limpeza dos grupos de emoção	71
6.1	Ferramentas de anotação: corte-e-costura e Vislcg3	72
6.2	Primeiro momento: Triagem dos lemas – passos 1, 2 e 3.....	77
6.3	Segundo momento: Informações adicionais – passos 4 e 5	80
6.4	Terceiro momento: Redefinição dos grupos – passos 6 e 7.....	83
7	Resultados: o Emocionário hoje	85
7.1	Resultados da ampliação e validação de padrões no Emocionário	85
7.2	Resultados da revisão do léxico do Emocionário	87
8	Considerações finais.....	92
9	Referências bibliográficas	95
Anexo 1	104
Anexo 2	111
Anexo 3	115
Anexo 4	122

Lista de Figuras

Figura 1 - Nuvem lexical do grupo Desespero na Literateca.....	17
Figura 2 - Abordagens de aprendizado de máquina e baseada em léxico para análise de sentimento. (Taboada, 2016:329)	26
Figura 3 - Definições de “gozar” pelos dicionários digitais Caldas Aulete e Priberam, respectivamente.....	47
Figura 4 - Distribuição de complementos do verbo “pedir” no corpus Obras.	53
Figura 5 - Resultados da busca pelo padrão léxico-gramatical “[pos="N.*" & func="SUBJ>"] "é" "um" "sentimento"”	55
Figura 6 - Distribuição das combinações de adjetivos para o lema “nervos”	69
Figura 7 - Distribuição por campo semântico do lema “satisfeito”	78
Figura 8 - Mostragem de linhas de concordância do lema “pedido” anotado como palavra de emoção.	78
Figura 9 - Distribuição semântica do lema “arrebatar” no AC/DC.	80
Figura 10 - Definição de “arrebatar” nos Dicionários Aulete Digital e Priberam, respectivamente.....	82
Figura 11 - Resultado de busca por “arrebatar é sentimento” no Google.	83

Lista de Tabelas

Tabela 1 - Léxicos de emoção e suas principais características.....	36
Tabela 2 - Comparação entre os léxicos para o português e o Emocionário.	37
Tabela 3 - Resultados da análise levando com conta a anotação do AC/DC.....	58
Tabela 4 - Panorama quantitativo da revisão do léxico do Emocionário. .	91

Lista de Quadros

Quadro 1 - Exemplos do lema “pena” com e sem sentido emocional.	15
Quadro 2 - Regras de anotação semântica de emoção para o lema “pena” no Emocionário.	15
Quadro 3 - Grupos de emoção no AC/DC.....	18
Quadro 4 - Exemplos de tipos de sentimento conforme os grupos de análise.....	57
Quadro 5 - Campos lexicais identificados no léxico de emoções.....	60
Quadro 6 - Exemplos de lemas com e sem erros de digitação.	61
Quadro 7 - Exemplos de lemas importados de outras línguas.	62
Quadro 8 - Alguns lemas que foram descartados ao longo da análise da cauda longa.....	63
Quadro 9 - Exemplos de lemas adicionados na anotação semântica com o padrão “sentimento de N”.....	65
Quadro 10 - Exemplos de regras positivas adicionadas ao Corte-e-Costura.....	76
Quadro 11 - Exemplos de regras negativas adicionadas ao Corte-e-Costura.....	76

1 Introdução

As discussões sobre definir o que é uma emoção ou um sentimento e definir quais emoções/sentimentos existem nas línguas perpassa várias áreas do conhecimento, como a antropologia, psicologia, e estudos da linguagem. O trabalho desta tese é, sobretudo, linguístico. O interesse recai sobre compreender como a língua portuguesa cristaliza a menção de emoção, sabendo-se que esta pesquisa não almeja estudar emoção em todos os vertentes. Não são analisadas emoções causadas no leitor, ou que estão por trás da escrita do autor ou que podem ser subentendidas pelas entrelinhas, por exemplo.

A pesquisa teve início com dois principais questionamentos – o que é emoção na língua portuguesa e quais emoções existem – para, a partir da detecção de palavras de emoção em grandes coleções de textos, criar um material que permita um amplo estudo da descrição de emoções em língua portuguesa. O ponto de partida para tentar respondê-los foi o Emocionário, projeto que estuda as emoções no português e que está descrito na seção 2.3 desta tese.

1.1 Motivação: estudo de emoções na língua

Emoção é um conceito que vem sendo estudado há muito tempo e a dificuldade quanto a uma definição precisa e/ou consensual do que seja emoção e/ou sentimento está presente, até hoje, em áreas diversas como psicologia, antropologia, filosofia e estudos da linguagem. Segundo o dicionário Houaiss Eletrônico 3.0, as definições de emoção são duas: “ato de deslocar, movimentar” e “agitação de sentimentos; abalo afetivo ou moral; turbação, comoção”. Para essa última, há uma definição complementar, da área da psicologia: “reação orgânica de intensidade e duração variáveis, geralmente acompanhada de alterações respiratórias, circulatórias etc. e de grande excitação mental”.

Nos estudos de psicologia, o pesquisador Arno Engelmann, ao final da década de 1970, já exprimia certa preocupação relacionada à linguagem: os nomes das emoções nas línguas humanas. Engelmann partiu da palavra “emoção” e outros conceitos considerados similares, como “sentimento”, “afeto”, “estado de ânimo” e “paixão” na língua portuguesa e em algumas outras (alemão, francês, inglês e italiano) para tentar definir o que se entende por “emoção” e “nomes de emoção” e como o conceito pode ser similar ou diferente dos outros.

De acordo com Engelmann (1978), os conceitos de “emoção” e “sentimento”, “afeto”, “estado de ânimo” e “paixão” já obtiveram as mais variadas definições, principalmente desde o início do século XX. Dentre tais tentativas de definição estão “um tipo particular de fenômeno de consciência”; “um tipo de comportamento”; e “tipos de fenômenos fisiológicos internos”, que poderiam ser considerados movimentos corporais, mudanças de caráter funcional no sistema nervoso, e fenômenos vasomotores.

Entre Aristóteles e os estoicos, as “emoções” eram consideradas “fenômenos discretos independentes”, que eram definidos a partir de “paixões básicas” (Engelmann, 1978:45). Segundo Engelmann, A partir do século XVIII, autores como Malebranche (1712) acreditavam que o amor era a “paixão-mãe”, de onde as consideradas “paixões básicas” se originavam, indicando que haveria certo grau de ligação entre fenômenos até então considerados independentes.

Até o século XX, as tentativas de definir o que são emoções e/ou sentimentos na psicologia variavam entre definições mais específicas, que consideravam “emoções” como “construtos teóricos que precisam ser inferidos para uma explicação adequada de certos comportamentos ou das relações entre situações e comportamentos” (Engelmann, 1978:43), e definições que quase alcançavam o ápice da generalização, pois caracterizavam “emoções” como sendo uma espécie de subárea da psicologia. A discussão em torno do nome “emoção” foi se tornando uma espécie de tabu, que não poderia ser substituído por um outro, com conotação científica, ou mesmo adotado em sua forma leiga, de forma a englobar conceitos pensados por cada autor, em cada momento dessa trajetória de tentativas de conceituação de “emoção” (Engelmann, 1978:79).

O desafio seguinte ao de definir o que é uma emoção é estabelecer quais emoções existem. A partir da segunda metade do século XX, alguns teóricos propõem a existência de emoções básicas (ou fundamentais) e comuns a todas

as culturas. Entre as teorias mais conhecidas estão a de Tomkins (1962, 1963), que desenvolveu a Teoria de Afetos (*Affect Theory*), que compreende nove emoções primárias divididas em positivas (*enjoyment/joy* e *interest/excitement*), neutras (*surprise/startle*) e negativas (*anger/rage*, *disgust*, *disgust*, *distress/anguish*, *fear/terror* e *shame/humiliation*); a de Ekman (1972, 1999), com sete emoções (*happiness*, *sadness*, *surprise*, *fear*, *anger*, *disgust* e *contempt*); e o modelo em forma de cone de Plutchik (1962, 1991), que postula oito emoções (*anger*, *fear*, *sadness*, *disgust*, *surprise*, *anticipation*, *trust* e *joy*).

O modelo de Ortony et al. (1988), também amplamente conhecido, não postula diretamente emoções básicas, mas sim sugere 22 tipos de emoção em uma estrutura que relaciona sujeito, objeto e ação a menções de emoção.

Uma abordagem para as emoções e sentimentos diametralmente oposta à ideia de emoções básicas é frequente nos estudos antropológicos: não há como se falar em universalidade, pelo contrário. As antropólogas Rezende e Coelho (2010) relatam a história de Laura Bohannan, antropóloga norte-americana que precisou se abrigar com uma família de uma tribo africana durante um período de tempestades. Ao compartilharem histórias, como era tradição na época de chuvas, Laura decide compartilhar a tragédia shakesperiana de Hamlet com a família. Várias dificuldades surgem por parte da família africana para seu entendimento, como a ideia de haver um fantasma, que na concepção daquela tribo não podem falar ou ter materialidade; a monogamia na relação matrimonial; a ordem hierárquica de sucessão; e, dentre outras, o ciúmes de Hamlet por sua mãe ter se casado novamente antes de se cumprirem os dois anos de luto, como era o costume. O objetivo de Rezende e Coelho ao relatarem essa passagem é expor os obstáculos à compreensão universal de sentimentos e relações sociais, argumentando que “sentimentos são tributários das relações sociais e do contexto cultural em que emergem” (2010:11) e que na visão das ciências sociais, “as emoções, embora situadas no corpo, têm com este uma relação que é permeada sempre por significados culturalmente e historicamente construídos” (2010:33).

Anteriormente, Catherine Lutz (1988) já partilhava da mesma visão, partindo da análise de expressão de emoção dos Ifaluk para tentar demonstrar que a experiência emocional é preeminentemente cultural e surge da convivência social. A autora corrobora com o pensamento de que tratar sobre emoções envolve falar também sobre sociedade, poder, política, padrões comportamentais (Lutz, 1988; Lutz & Abu-Lughod, 1990).

Na linguística, Wierzbicka (1999) critica expressamente todas as propostas de emoções básicas e, entre elas, a ontologia proposta por Ortony et al. (1988) (e, por extensão, as demais propostas de emoções básicas) por ser etnocêntrica e focada na língua inglesa. A autora acredita, ainda, que a palavra *emoção* pode estar muito engendrada à nossa língua nativa e às línguas predominantes no cenário acadêmico. Wierzbicka opta por adotar a palavra *emotion* em seu livro, porém não como uma ferramenta neutra e livre de problemas, mas como sinônimo de “sentimentos baseados em pensamentos” (1999:12). A autora enfatiza que não é possível discutir *emoção* sem usar palavras, por isso é fundamental estudarmos como elas são usadas em diferentes línguas e culturas.

Em uma publicação bem recente, Belinda Maia e Diana Santos (2018) fazem, também, um apanhado do conceito de *emoção* como foi descrito ao longo dos anos. As autoras, assim como Engelmann (1978), explicam que as tentativas de descrever “*emoção*” datam desde os filósofos gregos que viviam antes de Cristo e que usavam diferentes nomes para tipos de amor (*ágape*, *éros*, *ludus*, *pragma*, *philia*, *storge*). Isso se dá diferentemente do português, que usa adjetivos ou locuções para determinar o tipo de amor (“amor materno”, “amor romântico”, etc).

Visto que a discussão dos conceitos de *emoção* é extensa e problemática, as autoras decidiram focar seu estudo no motivo pelo qual essas tentativas de conceitualizar *emoções* podem ser tão diversas e contestáveis: elas dependem do viés de pesquisa adotado pelo estudioso. Como ilustram as autoras, se a revisão bibliográfica de *emoção* for feita à luz de alguns filósofos ou pesquisadores de psicologia e neurologia em trabalhos relacionando linguagem e computadores, áreas nas quais o estudo da *emoção* está relacionado à cognição, é muito comum encontrarmos a palavra “cognitivo”, por exemplo (2018:7).

Maia & Santos (2018), em sua discussão, dão mais ênfase à relação entre *emoção* e língua. Cada cultura e comunidade linguística experiencia *emoções* da sua maneira e a tradução de palavras relacionadas a *emoção* (*emotion*, *sensation*, *sentiment*, *feeling* e *mood*) geralmente não comporta todo o significado da palavra em sua língua fonte, pois podem ter diferentes conotações culturais nas culturas fonte e alvo. Acrescentam, ainda, que há diferenças culturais de sentido entre línguas, mesmo quando palavras-chave já padronizadas em uma língua são utilizadas para descrever uma *emoção*. As autoras estabelecem relações diretas entre *emoção* e as línguas. Os primeiros

paralelos estabelecidos foram com a sintaxe, por ser uma área mais objetiva da linguística. Pensando nos verbos, fazem referência à seguinte reflexão:

Teóricos da gramática de casos, que classificam os sujeitos e objetos gramaticais de acordo com seus papéis semânticos, concordaram de forma geral que nem a pessoa que sente a emoção, nem a causa aparente da emoção poderiam ser considerados atores ou agentes. Por exemplo, Halliday (1985) decidiu-se por casos com um *Experiencer* (para a pessoa que sentiu a emoção) e *Phenomenon* (para o que quer que tenha levado a tal emoção), e esses termos tornaram-se populares na literatura. Essas reflexões sugerem que, embora possamos aceitar um imperativo em uma frase como 'ame seu vizinho', é contra-intuitivo e bastante incomum acharmos exemplos ordenando que uma pessoa ame a outra. (MAIA & SANTOS, 2018:8)

E exemplificam como o *Phenomenon* é passível a ter subdivisões a partir da pesquisa feita por Maia (1994/1996) para sua tese de doutorado, na qual a autora usou onze subdivisões de *Phenomenon* relacionadas a sintaxe e ao contexto imediato para analisar exemplos de corpora com orações em inglês e português e demonstrar tanto os usos de palavras de emoção como quanto essas palavras estariam conectadas ao *Phenomenon* na escrita literária.

Ao relacionarem emoção e léxico, Maia & Santos (2018:10) alegam que, mesmo com todo o trabalho, identificação e hipóteses já alcançados por pesquisadores, ainda não há informação suficiente para dar conta do léxico de emoções nos usos reais da língua na área de Análise de Sentimento, tratada na seção 3.1 desta tese.

É com a perspectiva que nega a universalidade das emoções que a presente pesquisa se alinha, desconfiando de abordagens fundamentadas na ideia de emoções básicas, por um lado, e da tranquilidade relativa à sua tradução, por outro. É importante destacar que, na escrita desta tese, utilizo os termos "sentimento" e "emoção" como intercambiáveis. O interesse aqui está em detectar e caracterizar sentimentos e/ou emoções por meio de grandes coleções de textos em língua portuguesa, assim me afastando de discussões que possam, porventura, limitar as potencialidades da análise dos dados.

Para desenvolver esta pesquisa, minha proposta para descrever o campo semântico das palavras usadas para mencionar emoção e sentimento em grandes corpora de língua portuguesa ocorrerá a partir da busca por critérios para detectar e categorizar palavras de emoção a partir da exploração de padrões da língua e da aplicação de métodos que evidenciem a menção de emoção nos corpora, descritos nos capítulos 5 e 6 desta tese.

A partir da análise de dados produzidos pelos próprios falantes de português, esta pesquisa se desvia da discussão apresentada sobre o que são palavras de emoção, deslocando para os falantes de uma língua a

responsabilidade sobre a identificação de algo como sendo uma emoção ou sentimento. Especificamente, aproveito a disponibilidade de grandes corpora anotados para perguntar, à própria língua, como nós falantes usamos a língua portuguesa para descrever emoção.

1.2 Contexto

A partir de uma ampla varredura nos corpora do AC/DC (Santos e Bick, 2000)¹, projeto que reúne e disponibiliza gratuitamente recursos para pesquisa e corpora anotados em língua portuguesa, e atualmente conta com mais de um bilhão de palavras, utilizo como ponto de partida o Emocionário. O Emocionário teve início em 2015 e ganhou esse nome em 2019. Ele documenta a anotação semântica de emoção do AC/DC: é um projeto de anotação de emoções e também um léxico de emoção criado a partir da observação de corpus e buscas em dicionários. Sua atual configuração é resultado da terceira fase de desenvolvimento, que ocorreu durante a confecção desta tese, e o processo é descrito detalhadamente na seção 2.3. O Emocionário contém atualmente 26 grupos de palavras usadas para descrever emoção na língua portuguesa e faz parte da Gramateca, que viabiliza o estudo da gramática da língua portuguesa e é tratada na seção 2.2 desta tese.

1.3 Objetivos

Esta tese tem como objetivo geral criar subsídios para a anotação do campo semântico das emoções na língua portuguesa, cuja metodologia poderá servir para outras línguas futuramente, e como um dos objetivos específicos, contribuir para que o acervo do AC/DC seja um material de referência na área de anotação de emoções nos estudos com a língua portuguesa. Isto porque, apesar de já dispormos de léxicos de palavras de emoções e sentimentos para o português (como mostra a seção 3.2 desta tese), apenas uma lista de palavras

¹ Ele será tratado de maneira mais detalhada na seção 2.2 desta tese.

² <https://www.linguateca.pt/>

³ As regras do Emocionário estão disponíveis ao público e podem ser encontradas em

pode ser insuficiente quando se trata de emoção, como trata o capítulo 3. Além disso, até a finalização desta tese, não tivemos conhecimento de projetos de anotação de palavras de emoção para a língua portuguesa, o que reforça sua relevância.

Cada língua, em particular, tem uma série de valores culturais e sociais que se encontram no léxico e na forma como os falantes se expressam através dela, e que estão em constantes mudanças ao longo do tempo (Boddice, 2018). A opção de utilizar grandes corpora valoriza a análise feita a partir de exemplos reais produzidos pelos falantes da própria língua, ainda que representem apenas uma amostra, por maiores que sejam esses corpora.

Outros objetivos específicos são:

- (i) Descrever o campo semântico das palavras usadas para indicar emoção em grandes corpora de língua portuguesa;
- (ii) Validar os grupos de emoção do léxico do Emocionário para classificar as palavras que descrevem emoção; e
- (iii) Apresentar critérios para atribuir emocionalidade a uma palavra, partindo de explorações em corpora com padrões que evidenciem a menção de emoção na língua portuguesa;
- (iv) Reavaliar, revisar e reorganizar o léxico e anotações semânticas do Emocionário, criando um material ainda mais refinado e abrangente.

Esta tese está organizada da seguinte maneira: o presente capítulo abordou a motivação desta pesquisa e trouxe a discussão quanto aos desafios envolvidos na análise de definição e descrição de emoção, que atravessam diferentes áreas de estudo e estão presentes não apenas na língua portuguesa. Também explicitou o contexto da pesquisa e descreveu os objetivos gerais e específicos que espero alcançar ao longo dos capítulos seguintes.

O segundo capítulo se ocupa do enquadramento teórico-metodológico da tese. Nele, estão descritos estudos linguísticos com corpus, o AC/DC, a Linateca e Gramateca, pontos de partida para o trabalho desta tese. Também são apresentadas as duas frentes do Emocionário – o projeto de anotação das emoções e o léxico de emoções da Linateca. A seguir, explico brevemente as Humanidades Digitais e a Leitura Distante, e o fim do capítulo apresenta resumos de pesquisas que dialogam com corpus, leitura distante e análise de menção de emoção.

No terceiro capítulo, descrevo a tarefa da Análise de Sentimento e explico por que esta pesquisa não se insere na área e, em seguida, descrevo detalhadamente alguns dos léxicos que já existem para a língua portuguesa. Ao

final desse capítulo, listo os motivos pelos quais esses léxicos e o Emocionário não podem ser comparados.

O quarto capítulo traça um panorama das características e desafios da análise de descrição de emoção a partir de casos encontrados ao longo da análise de lemas que fazem parte do léxico do Emocionário. As seções desse capítulo tentam seguir caminhos dentro da observação da língua para confirmar se uma palavra ou expressão é usada para descrever emoção; em que casos ela pode ser considerada descrição de emoção; e a quais grupos do Emocionário ela pode vir a pertencer.

Os capítulos cinco e seis da tese detalham as metodologias desenvolvidas ao longo desta pesquisa, que materializam a fase atual de construção do Emocionário. O quinto capítulo trata das explorações em corpora. Mais especificamente, descrevo como ocorreu a pesquisa pelo padrão “sentimento de N” e detalho o tratamento dos lemas com menos de três ocorrências no corpus (cauda longa); e a pesquisa por conexões entre menção de emoção através de expressões, resultante da exploração da anotação de palavras do corpo humano do projeto Esqueleto.

No sexto capítulo, descrevo de forma detalhada a metodologia desenvolvida para revisão do Emocionário, ilustrando cada passo a partir de exemplos retirados do próprio corpus no processo de aplicação do método. No final do capítulo, comento o produto da revisão em termos quantitativos e qualitativos.

O sétimo capítulo reúne os resultados das metodologias expostas nos capítulos cinco e seis; e o oitavo capítulo aborda as considerações finais, em que analiso quais dos objetivos foram atingidos; menciono possíveis limitações e desafios na construção desta pesquisa; listo as conclusões obtidas após as explorações expostas ao longo dos capítulos anteriores e contribuições que acredito que esta pesquisa possa oferecer a pesquisadores da linguagem ou áreas que tenham interesse pelo processamento de linguagem natural. Ao final, proponho desdobramentos futuros como aprofundamento ou produção de novas pesquisas a partir do material aqui apresentado.

2 Áreas de pesquisa: enquadramento teórico metodológico

Este capítulo é dedicado às áreas de pesquisa nas quais esta tese se insere. A seção 2.1 descreve os estudos linguísticos com corpus eletrônico; seus possíveis usos práticos; discorre sobre a importância da escolha do corpus estar em consonância com os objetivos da pesquisa a ser realizada. Além disso, elenca os possíveis tipos de anotação que podem ser feitas e as vantagens de se trabalhar com corpus. A seção 2.2 apresenta o AC/DC, a Linguateca, a Gramateca e as ferramentas de anotação e validação disponíveis. A seção 2.3 descreve as duas frentes do Emocionário – o projeto de anotação das emoções e o léxico de emoções da Linguateca –, suas principais características, perpassando as fases de desenvolvimento desde a sua criação até o atual momento. A seguir, a seção 2.4 explica brevemente as Humanidades Digitais e discorre em mais detalhes sobre a Leitura Distante, e a seção 2.5 finaliza o capítulo ilustrando, a partir das pesquisas de Klinger et al. (2016), Grayson et al. (2016) e Heuser et al. (2016) com análises de obras literárias que demonstram como a sinergia entre corpus, leitura distante e análise de menção de emoção é possível e muito proveitosa para o desenvolvimento de pesquisas e, ao mesmo tempo, como pesquisar emoção é desafiador.

2.1 Estudos linguísticos com corpus

A área de Estudos Linguísticos com Corpus se popularizou na década de 1990 e se interessa em estudar a língua através de exemplos reais de uso. Um corpus, como defendeu Santos (2008:45), é “uma coleção classificada de objectos linguísticos para uso em Processamento de Linguagem Natural / Linguística Computacional / Linguística”, em que, segundo a autora, “uso” se refere ao objetivo da pesquisa, seja ela um estudo, uma avaliação, uma medição; “objectos linguísticos” são os dados, podendo ser textos de variados gêneros textuais, frases, palavras; e “coleção classificada” pode incluir, por exemplo, as categorias a serem consideradas na pesquisa; tipos de erros, tipos

de traduções; e constituintes dos dados, como análise semântica, sintática, fonológica ou discursiva.

As potencialidades dos estudos com corpus envolvem, dentre outras, auxiliar pesquisas de cunho lexicográfico e terminológico; descrição de língua materna; processamento de linguagem natural (PLN); e ensino de língua estrangeira e tradução. A abordagem envolvendo corpus permite estudar e descrever a linguagem; criar e testar hipóteses; e observar padrões mas também fenômenos raros presentes naquela amostra representativa do objeto da análise.

Um questionamento relativamente comum no uso de corpus em pesquisas é o tamanho do corpus a ser utilizado. O corpus deve funcionar como uma amostra do objeto da análise, podendo ser sobre um (ou mais) aspecto da língua ou de uma variedade linguística. Para detectar padrões em um corpus com textos de uma determinada variedade linguística, naturalmente quanto maior for a amostra, mais casos incomuns serão encontrados, bem como maiores serão as chances de serem encontradas frequências e/ou padrões que possam exemplificar teorias já existentes na língua, não havendo limite de tamanho para a construção desse corpus.

No entanto, os corpora cujos objetivos incluem analisar um autor, por exemplo, acabam por compilar uma quantidade fixa de palavras, que não necessariamente resultará em um corpus volumoso, mas servirá ao objetivo. Se o pesquisador tem interesse em analisar a obra de Machado de Assis, por exemplo, ele poderá compilar um corpus que abranja toda a obra do autor. O mesmo ocorreria com um pesquisador interessado em fazer um estudo sobre publicações de um período específico de um jornal – ele seria capaz de coletar todos os artigos publicados, porque trata-se de um número limitado.

Independente do tamanho do corpus, é importante que ele seja manipulado através de softwares de análise linguística e que, preferencialmente, contenha informações linguísticas associadas a ela, sendo metadados com informações como o autor do texto, gênero textual e data de publicação, ou anotações em um nível linguístico referentes a classes de palavras, sintaxe e campo semântico, por exemplo. A anotação linguística é útil pois enriquece as possibilidades de busca e possibilita que esses dados sejam estatizados e comparados.

No caso da anotação em corpus, o pesquisador pode usar uma lista de categorias prévias para conduzir a anotação. Acontece que, ao analisar o corpus, é muito comum que ele se depare com fenômenos linguísticos que se encaixem em mais de uma ou até em nenhuma das categorias escolhidas

previamente. Para Baker (2010:19), é importante o pesquisador se conscientizar de que a consistência é a chave da categorização, reconhecendo que toda categorização é “impressionista” e incapaz de dar conta das especificidades das línguas. Segundo o autor, o pesquisador deve, em seu processo de categorização, decidir entre modificar as categorias escolhidas antes da análise; inventar categorias a partir de combinações de palavras (como por exemplo “amocidade” – uma combinação entre “amor” e “felicidade”); ou encaixar os resultados nas categorias escolhidas, mesmo sabendo que qualquer uma das opções não dará conta da totalidade do significado. Nesta tese, opto por criar novas categorias além das preexistentes no Emocionário (seção 2.3 e capítulo 5) e inserir as palavras que cabem em mais de um grupo nos diferentes grupos que combinam com seus sentidos emocionais.

Entre as maiores vantagens de estudar a partir de corpora está a possibilidade de se analisar e processar grandes quantidades de texto em um curto período de tempo, pelo fato de esse corpus ser processado por uma ferramenta computacional. É importante destacar que o corpus oferece os dados, que só fazem sentido quando o pesquisador os maneja e interpreta dentro das teorias de modo a direcionar seu estudo de acordo com seus objetivos.

Na análise quantitativa promovida pela linguística que utiliza corpora, os números são apenas o ponto de partida de uma análise interpretativa. Os estudos com corpus facilitam o acúmulo de métodos de análise simultâneos; proporcionam a detecção de padrões linguísticos e a frequência com que aparecem; e possibilitam a análise de grandes quantidades de texto.

2.2 AC/DC, Linguateca e Gramateca

A Linguateca² (Santos, 2002; Santos et al., 2004; Santos e Costa, 2005) é um centro de recursos distribuído para o processamento computacional da língua portuguesa cujo objetivo geral é oferecer meios para o estudo da língua através do processamento de dados. Ela teve início em 1998, mas o nome atual só foi adotado em 2002. Suas principais linhas de trabalho são catalogar a língua portuguesa através de processamento de linguagem natural e da

² <https://www.linguateca.pt/>

linguística computacional; criar e distribuir recursos para a análise da língua de forma colaborativa; e avaliar tarefas/programas em língua portuguesa buscando envolver a comunidade científica. Vale ressaltar que um eixo de muita importância em todo esse trabalho sempre foi disponibilizar todos os recursos de forma pública e gratuita para a comunidade linguística – o interesse da Linguateca era principalmente servir aos pesquisadores distribuindo recursos para o processamento do português.

Dentre as atividades realizadas pela Linguateca estão a formação especializada de pesquisadores em gestão, criação, disseminação e avaliação de recursos; a supervisão e manutenção de serviços de repositório, distribuição, catalogação, contato e comunicação entre produtores de conteúdo e interessados nos serviços oferecidos na página; o desenvolvimento de recursos públicos para avaliação ou calibragem; e a organização de avaliações conjuntas em torno de áreas chave nos estudos da linguagem. Além disso, a Linguateca comporta projetos como o COMPARA e o PANTERA, que são corpora paralelos português-inglês e português-norueguês, respectivamente; o HAREM, recurso de avaliação para Reconhecimento de Entidades Mencionadas em Português; o Rêve, ferramenta de revisão da anotação para avaliar concordância entre anotadores e melhorar a anotação já existente na plataforma.

O AC/DC (Santos e Bick, 2000), que significa Acesso a Corpora / Disponibilização de Corpora, foi um dos primeiros projetos desenvolvidos pela Linguateca e teve início em 1999 como uma tentativa de agrupar recursos e corpora de língua portuguesa disponíveis no intuito de facilitar a comparação e exploração desse material. O resultado seria uma interface com uma grande coleção de dados disponíveis e com acesso ao Open CWB, uma ferramenta de alta qualidade para interrogação de corpora.

O projeto pretendia identificar e adequar os formatos dos recursos que já se encontravam disponíveis para a língua portuguesa para serem servidos pela Linguateca. Em 1999, contava com cinco corpora diferentes. Até o final de 2021, quarenta corpora diferentes faziam parte do AC/DC. Já com o objetivo de enriquecer e aprimorar a anotação morfológica, sintática e morfossintática, foi utilizado o PALAVRAS (Bick, 2000), um *parser* que dá conta da língua portuguesa e está descrito na seção 2.2 desta tese.

A Gramateca é uma tentativa de criar um laboratório sobre o AC/DC e outros corpora. Segundo Santos (2014), a Gramateca “é mais uma infraestrutura para fazer gramática(s) ou estudos de gramática do que algo que produza mais uma nova gramática”. O ambiente da Gramateca tem o intuito de

promover o máximo de interação possível entre áreas, como a linguística, o processamento de linguagem natural e as Humanidades Digitais, por exemplo, e também permite uma variedade de estudos linguísticos, não se limitando apenas à sintaxe. Isto é, mesmo sabendo que a gramática tradicionalmente se ocupa da morfossintaxe, a Gramateca também é composta por estudos de semântica e léxico, não se limitando a áreas e definições tradicionais dos estudos das línguas. Algumas das possibilidades de trabalho com a Gramateca são acompanhar o uso e a evolução de um dado campo semântico através do tempo; diferenças entre obras literárias no uso de palavras de cor e/ou de descrição de emoção; o uso da segunda pessoa em diferentes gêneros; a distribuição de nomes próprios por tema; além de estudos contrastivos por meio do COMPARA, CorTrad e PoNTE, corpora paralelos que podem fornecer pistas para que sejam comparadas com achados dos corpora monolíngues.

Por fim, o diferencial da Gramateca é seu potencial de combinar e integrar áreas e temas, qualitativa e quantitativamente, e não só descrevendo, mas também explorando dados, testando hipóteses e avaliando achados. O único pré-requisito para publicar algo associado à Gramateca é que os dados e resultados, além das conclusões, estejam divulgados por completo. Isto porque, assim, pode haver uma comunicação direta (e com o mínimo possível de interferências) entre dados, pesquisas e pesquisadores.

2.3 O Emocionário

O projeto de anotação semântica e léxico de emoções do AC/DC começou a ser desenvolvido no âmbito da Gramateca por Cristina Mota e Diana Santos (2015; também Maia e Santos, 2012; Mota, 2013; Santos et al., 2021). Sua primeira fase de desenvolvimento teve os objetivos de estudar como falantes da língua portuguesa usam palavras de emoção e de aumentar a informação semântica nos corpora da Linguateca.

À época, o léxico foi construído a partir de dicionários e ontologias lexicais e revisado manualmente (Mota e Santos, 2015). A construção do léxico, como ressaltado pelas próprias autoras, foi um pontapé para a documentação e anotação da menção de emoção que precisaria de posterior revisão e reavaliação. Nessa primeira fase do trabalho, uma grande quantidade de

informação foi agrupada de forma automática e essa informação foi dividida em dezenove grupos de palavras de emoção, e cada palavra foi caracterizada entre “emomin”, se a maioria de seus usos no corpus descrevesse emoção, ou “emomax”, quando seu principal uso não fosse para descrever emoção. Nesse ponto, diversos problemas decorrentes da detecção automática foram identificados (ver capítulo 4 desta tese para desafios da análise de emoção), como de que modo anotar palavras polissêmicas que descrevem emoção apenas em certos contextos (por exemplo, “pena”) e palavras usadas em expressões idiomáticas (por exemplo, “rua da amargura”). O objetivo principal da primeira fase, que foi cumprido, era oferecer um conjunto de dados à comunidade linguística como uma tentativa de anotar emoções.

A segunda fase do projeto (Santos et al., 2021) levou à criação de quatro novos grupos para o léxico, totalizando 24 grupos de emoção que comportam aproximadamente 2.800 lemas diferentes e cerca de 8.000 formas de palavras que são usadas na língua portuguesa para fazer referência à emoção. Extinguiu também as subdivisões da anotação das palavras de emoção em “emomin” e “emomax”. As regras para a anotação semântica de emoção são feitas através do Vislcg3 e do corte-e-costura (que serão descritos na seção 6.1 desta tese). Há a possibilidade de mudança, adição e remoção das regras a qualquer momento no processo de observação da língua, que está em constante mudança – por isso ele é um projeto de anotação das emoções.

Com base nas regras³ do projeto de anotação do Emocionário é feita a anotação nos corpora. Essas regras podem ser mais simples, indicando que a palavra é uma emoção, ou mais complexas – que acabam ocorrendo em maioria – quando a palavra indica emoção em certos contextos, como mostra o Quadro 1, com exemplos do lema “pena”.

Exemplos do lema “pena” com sentido emocional:

Expressão de busca: [lema="pena" & sema=".*emo.*"]

- par=ext161210-opi-96b-1: É **pena** que o espírito «sem limites» se limite aos desporto e a outras actividades recreativas.
- <p>: Revelou que ficava com **pena** da mãe."
- par=FSP940117-141-1145: Tenho **pena** do Marcelo Paiva, que escreve ótimas matérias no meio de tanta estupidez.
- <p>: Agora o que sentia era **pena** mesclada de curiosidade.

³ As regras do Emocionário estão disponíveis ao público e podem ser encontradas em https://www.linguateca.pt/aceso/corpos/obras/regras_emodizer.utf8.txt; https://www.linguateca.pt/aceso/corpos/obras/todas_regras.txt; e https://www.linguateca.pt/aceso/corpos/<nome do corpo>/todas_regras_excl.txt (para as regras específicas para cada corpus).

Exemplos do lema “pena” sem sentido emocional:

Expressão de busca: [lema="pena" & sema="0"]

- par=ext547877-soc-98a-1: Depois de uma semana às voltas com a Educação em Portugal, o Presidente da República disse ontem que vale a **pena** fazer política «na sua forma mais nobre», no terreno.
- par=PUBLICO-19950615-068-562: As **penas** previstas são as mesmas do caso anterior, com as mesmas atenuantes.
- par=FSP950820-206-2963: Jamais uma pessoa que tem dinheiro vai morrer na **pena** de morte.
- <p>: A exuberância se concentra nos pêlos e **penas** dos animais, no figurino com colagens exóticas.

Quadro 1 - Exemplos do lema “pena” com e sem sentido emocional.

Para a separação dos casos com e sem sentido emocional, que estão no Quadro 1, foram feitas regras gerais, que se aplicam a todos os corpora, e positivas, o que significa que o lema “pena” originalmente não carrega uma anotação semântica e só ganha a etiqueta de emoção nos casos incluídos nas regras do Quadro 2. Em alguns casos, “pena” aparece referido como *word* em vez de *lema*, para anotar as entradas de “pena” exatamente na forma que está na expressão de busca – singular – e, em um dos casos, com letra maiúscula.

Regras de anotação semântica de emoção para o lema “pena” no Emocionário:**“Ter pena”:**

[lema="ter"] a:[word="pena"] >> a:[sema="emo:pena"]

“Ter muita pena / Ter tanta pena / Ter imensa pena”:

[lema="ter"] [word="muita|tanta|imensa"] a:[word="pena"] >> a:[sema="emo:pena"]

“Sem pena / Com pena”:

[word="sem|com"] a:[word="pena"] >> a:[sema="emo:pena"]

“Com muita pena / Com tanta pena / Com imensa pena”:

[word="com"] [word="muita|tanta|imensa"] a:[word="pena"] >> a:[sema="emo:pena"]

“Ser pena”:

[lema="ser"] a:[word="pena"] >> a:[sema="emo:pena"]

“Que pena” seguida de verbo no infinitivo:

[word="que"] a:[word="pena"] [temcagr=".*INF.*"] >> a:[sema="emo:pena"]

“Que é uma pena / Que foi uma pena”:

[word="que"] [word="é|foi"] [word="uma"] a:[word="pena"] >> a:[sema="emo:pena"]

“Pena que”:

a:[word="Pena"] [word="que"] >> a:[sema="emo:pena"]

Quadro 2 - Regras de anotação semântica de emoção para o lema “pena” no Emocionário.

A anotação semântica tem como principal característica a generalização de casos, que são agrupados e etiquetados. A criação ou organização de campos semânticos é sempre uma tentativa de evidenciar o sentido das palavras e funciona como uma sistematização do processo de categorização que o falante faz naturalmente na construção da língua. Aqui me refiro à categorização pela visão de John Ellis (1993:29), como um processo que resume experiências e situações em grupos finitos que se materializam nos léxicos das línguas. De

acordo com o autor, seria impossível nomearmos e comunicarmos cada experiência vivida, pois elas são únicas, infinitas e têm lugar no passado:

O problema é que os fatos da experiência são infinitamente variáveis: nenhuma situação é exatamente igual a outra. A ideia de nomear cada situação diferente é impossível. Primeiro, o nome não faria sentido para um falante B, a não ser que ele tivesse presenciado todo o acontecimento. [...] Em segundo lugar, a quantidade de nomes seria rapidamente tão grande que nenhum dos falantes conseguiria lembrar sua referência. Mas, em terceiro lugar, e o motivo mais esclarecedor, é que a língua não teria utilidade. Seria limitada a relembrar situações passadas, e principalmente nada seria dito sobre essas situações. (Ellis, 1993:28)

Logo, ao invés de focarmos em nomearmos as diferenças entre cada vivência, ao categorizar o foco recai sobre distribuir uma grande quantidade de casos a um número muito menor de grupos que façam sentido para o propósito dos falantes daquela comunidade linguística, seja encontrando padrões a partir de experiências abstratas ou buscando simplificar complexidades a fim de promover a comunicação.

O léxico de uma língua é suscetível a variação, mudança e invenções, o que significa que toda tentativa de agrupar palavras envolve um trabalho complexo e parcial, por mais abrangente que seja. Os léxicos das línguas ganham estrutura pelo processo de categorização e se substancializam nos campos lexicais. Logo, os grupos do Emocionário ao mesmo tempo diferenciam e igualam palavras que descrevem emoções usadas pelo falante do português, para assim ser possível lidar com milhares de itens lexicais diferentes e analisar os dados seguindo esse critério da presença de emoções.

A imensa discussão sobre a própria palavra “emoção” (ou “sentimento”) ilustra a dificuldade em nomear grupos ou tipos de emoção, em oposição à facilidade na compreensão das polaridades (positiva, negativa ou neutra). O que acontece no caso dos títulos de cada grupo do léxico de emoções é que não são, em si, um aspecto muito importante, pois são usados apenas como formas de organização e poderiam ser até palavras inventadas. Por isso, no Emocionário, a escolha foi por nomear os grupos pelos próprios nomes de algumas emoções/sentimentos, na esperança de serem mais informativos do que um código ou uma palavra inventada.

A desvantagem dessa opção é que nem sempre a palavra escolhida para o título dá conta de todos os sentidos envolvidos no grupo. No grupo Desespero, por exemplo, estão contidos os sentidos equivalentes a “aflição”, “decepção”, “desespero”, “desesperança”, “frustração”, “falta de confiança” e “desconfiança”.

A Figura 1 mostra uma nuvem lexical do grupo Desespero no corpus Literateca em novembro de 2019.



Figura 1 - Nuvem lexical do grupo Desespero na Literateca.

De modo a esclarecer o usuário da Linguateca, as listas de palavras que compõem os grupos do Emocionário são acompanhadas da explicação dos possíveis sentidos em cada grupo de emoção, reproduzidas a seguir, no Quadro 3. Essas explicações foram produzidas em 2019 a partir de meus primeiros contatos com os grupos, em uma tentativa de auto-organização. Essas definições também estão disponíveis na página da Gramateca⁴. Destaco aqui o fato de o grupo Admirar ser o único grupo cujo título é um verbo pois, à época da definição dos grupos e seus respectivos títulos na idealização do projeto de estudo das emoções, a decisão foi feita por nomeá-los com a forma mais frequente de cada grupo, o que coincidentemente resultou em uma falta de paralelismo sintático.

Grupos de emoção no AC/DC:

Admirar: engloba sentidos como extasiar-se diante de algo, experimentar sentimento, respeito, simpatia, amor; ter ou sentir admiração recíproca ou por si mesmo;

Alívio: engloba sentidos como diminuição de dor, peso, trabalho; desopressão, desafogo; sentir-se tranquilo; experimentar consolo, refrigério;

Amor: engloba sentidos como desejar o bem do outro ou de alguma coisa; sentimento de dedicação absoluta de um ser a outro ser ou a uma coisa; atração por outra pessoa, também de caráter sexual; inclinação, apego ao que desperta prazer ou empatia; cuidado, zelo, dedicação;

Coragem: engloba sentidos como força moral e perseverança no enfrentamento de situações emocionalmente difíceis; disposição para suportar ou superar problemas; resolução, franqueza, desembaraço; determinação, constância, perseverança, ousadia;

⁴ Disponível em: <https://www.linguateca.pt/Gramateca/Emocionario.html>.

<p>atitude desaforada ou desavergonhada; vaidade exagerada, jactância, soberba;</p> <p>Desejo: engloba sentidos como vontade, anseio; cobiça, ambição por alguma coisa; aspirar ao que não possui; apetite sexual;</p> <p>Desespero: engloba sentidos como estado de espírito ou sofrimento daquele que passa por inúmeras dificuldades e aflições e não tem como superá-las ou acredita que não o possa fazer; desesperança; sofrimento moral extremo, misto de aflição, angústia, descontrole e tormento;</p> <p>Esperança: engloba sentidos como a expectativa otimista da realização daquilo que se almeja, mesmo que pouco provável, ilusório, vão;</p> <p>Felicidade: engloba sentidos como bom êxito, sucesso; contentamento; alegria;</p> <p>Fúria: engloba sentidos como agitação violenta; exaltação de ânimo;</p> <p>Genérica: engloba casos que fazem menção a uma emoção não identificada ou não especificada;</p> <p>Gratidão: engloba sentidos como reconhecimento; agradecimento;</p> <p>Humildade: engloba sentidos como modéstia; simplicidade; consciência das próprias limitações;</p> <p>Infelicidade: engloba sentidos opostos ao do grupo "felicidade";</p> <p>Ingratidão: engloba sentidos opostos ao do grupo "gratidão";</p> <p>Insatisfação: engloba sentidos opostos ao do grupo "satisfação";</p> <p>Inveja: engloba sentidos como o desgosto ou pesar pela felicidade de outros; forte desejo de possuir o que é de outros; incapacidade de aceitar que a pessoa querida ou amada goste de outros;</p> <p>Medo: engloba sentidos como uma grande inquietação frente a uma ameaça ou perigo, físico ou abstrato, real ou imaginário;</p> <p>Ódio: engloba sentidos como aversão a outra pessoa, situação ou coisa; rancor ou inimizade que impelem alguém a causar ou desejar o mal;</p> <p>Orgulho: engloba sentidos com conotação positiva, como o sentimento de dignidade pessoal, e com conotação negativa, como um conceito elevado de si próprio; soberba.</p> <p>Pena: engloba sentidos como padecimento; compaixão;</p> <p>Satisfação: engloba sentidos como contentamento; prazer após a realização de algo ou um acontecimento proposital ou acidental;</p> <p>Saudade: engloba sentidos como uma "lembrança nostálgica e, ao mesmo tempo, suave, de pessoas ou coisas distantes ou extintas, acompanhada do desejo de tornar a vê-las ou possuí-las";</p> <p>Surpresa: engloba sentidos como um espanto provocado por um acontecimento inesperado.</p> <p>Vergonha: engloba sentidos como insegurança provocada pelo excesso de exposição; desonra ou humilhação frente a outros.</p>

Quadro 3 - Grupos de emoção no AC/DC.

O projeto de anotação do Emocionário viabiliza a construção de um léxico baseado na análise de dados produzidos pelos próprios falantes do português. É também o projeto de anotação das emoções que permite que o léxico do Emocionário possa ser enriquecido e aprimorado a partir do estudo da descrição de emoções na língua portuguesa. Por existirem palavras que remetem a emoção em apenas alguns contextos, construir um léxico com palavras que sejam apenas usadas para descrever emoção limitaria a riqueza lexical do projeto. Por isso, a característica de maior destaque do Emocionário é sua flexibilidade na inserção de regras que garantam que as palavras que compõem

o léxico sejam consideradas em seu sentido emocional. De um modo geral, são três as principais características do léxico do Emocionário:

- (i) a possibilidade de um mesmo lema fazer parte de mais de um grupo de emoção, trabalhando de forma complementar segundo os sentidos das palavras ou indicando na anotação os casos que pertencem a cada grupo, como indicam Mota e Santos (2015) em um estudo sobre “admiração” com sentido de “espanto” ou “respeito”. Os Exemplos 1 e 2 abaixo mostram o lema “admirar” no corpus Literateca, primeiro anotado como parte do grupo Surpresa e, em seguida, do grupo Admiração:

Exemplo 1: <p>: Ergue-te, filho; Eu fiz o meu dever: não te acostumes A **admirar** com espanto uma acção boa.

Exemplo 2: par=CCB-Anátoma-191: O estudante de dezoito anos era **admirado** como um prodígio de ciência infusa e de proféticas e indeterminadas aspirações para a ciência .

- (ii) a abrangência semântica de cada grupo, que faz com que lemas como “aflito” e “atormentado” pertençam ao mesmo grupo (neste caso, o grupo Desespero), ainda que estejam distantes um do outro se considerados “níveis de desespero”, e também aparecem ilustrados nos Exemplos 3 a 6;

Exemplo 3: -- Sem encontrarem o xerife negro e a sua bem-amada, os **aflitos** emparedados que passaram séculos a soluçar amarguras por horas mortas da noite! -- gracejou o Vasconcelos -- Pois muitos vos agradeço o conto da velhita, vossa criada, e daqui o vou já dizer ao meu amigo Vasco Eanes para o convencer de que sonhou acordado .

Exemplo 4: Passados momentos, entrou Rosa no quarto, e, ao ver aqueles rostos **aflitos** e aqueles corações entregues à mais inconsolável dor, sentiu também as lágrimas caírem-lhe uma a uma pelas faces, e, como atormentada por um triste pressentimento, correu para junto do leito, e com a voz ansiada chamou Fernando .

Exemplo 5: <p>: Por isso, um homem permanentemente **atormentado**, «de 'pírito inquieto e insatisfeito consigo mesmo e com quase tudo que vê à volta de si», como reconheceu em seu discurso de posse na Academia Brasileira de Letras, em 1968 .

Exemplo 6: <p>: Em um mesmo ano, Iberê produz essas figuras **atormentadas**, prestes a se diluírem na matéria pastosa da pintura, mas ainda agarrando-se a contornos que buscam o traço simples do desenho esquemático.

E (iii) a inclusão de palavras independentemente de sua frequência no corpus, sinalizando que carregam emoção em algum contexto, como os lemas “choque”, com 2.768 entradas que descrevem emoção de um total de 39.906 entradas deste lema no AC/DC, e “atônito” (ou “atónito”), com um total de 1.857 entradas no corpus, sendo todas correspondentes à descrição de emoção. Ambos os lemas “choque” e “atônito” (ou “atónito”) são parte do grupo

“Surpresa”, mesmo que em proporções quantitativas distintas no AC/DC e no léxico do Emocionário os dois apareçam listados porque descrevem emoção. Há, inclusive, lemas com número de entradas ainda menor que carregam emoção, como o lema “jubilação”, do grupo Satisfação, que existe no AC/DC com 176 entradas, sendo apenas doze referentes à emoção⁵.

Apenas em 2019, foi atribuído o nome de Emocionário à combinação do projeto de anotação semântica das emoções, em um sentido mais amplo, e do léxico das emoções, em um sentido mais estrito. O resultado final encontrado pelo usuário da Linguateca é tanto a listagem de lemas dos grupos de emoção no léxico na Gramateca como aquele obtido nas buscas por distribuição semântica nos corpora do AC/DC, produto das regras de anotação semântica do projeto, como o próprio corpus do AC/DC que passa a ser um corpus anotado com informação semântica de emoção.

Em sua terceira fase, iniciada a partir da pesquisa desta tese, o Emocionário vem sendo reavaliado, revisado, reorganizado e ampliado, com o objetivo de disponibilizar um material de anotação ainda mais refinado e preciso à comunidade linguística. Uma das primeiras opções consideradas para a reorganização dos grupos foi a junção de grupos considerados opostos em sentido, como Amor/Ódio, Felicidade/Infelicidade, Satisfação/Insatisfação, Esperança/Desespero. Há, no entanto, dois argumentos principais para não optarmos pela junção: o primeiro é o “empobrecimento” da busca, pois minimizaria consideravelmente a possibilidade de busca no corpus e o pesquisador não teria a possibilidade de procurar separadamente por cada um dos grupos no texto. Explicando: se o pesquisador estiver interessado em procurar menções de Satisfação no corpus, inevitavelmente encontrará menções de Insatisfação na mesma busca, o que pode não ser relevante para sua pesquisa. Ao mantermos os grupos separados, portanto, os aspectos da busca continuam mais refinados e, caso deseje, o pesquisador poderá juntar as análises de grupos diferentes.

O segundo argumento é que há uma espécie de “névoa semântica” pairando sobre alguns dos grupos, algo que chamou atenção durante a análise dos grupos do léxico de emoções considerando a junção. Isto é, alguns grupos não teriam opostos diretos. Uma das opções a se considerar seria a junção dos grupos Orgulho, Vergonha e Humildade. Vale ressaltar que a polaridade não foi um critério definidor para todos os grupos: os pares “coragem”/“atrevido” e

⁵ Na versão 7.7 do AC/DC; último acesso em dezembro de 2021.

“medroso”/“temente”, por exemplo, fazem parte dos mesmos grupos, Coragem e Medo, respectivamente. Ainda sim, foi levada em consideração, visto que seria impossível desassociar a orientação semântica positiva ou negativa das palavras adicionadas ao Emocionário. Logo, alguns grupos acabam sendo, sim, opostos, como Amor/Ódio e Felicidade/Infelicidade. No entanto, em relação a polaridade, “orgulho” contraposto com “vergonha” é positivo, ao passo que contraposto com “humildade”, é negativo. Assim, não seria possível olhar para os grupos com as noções de gradação positiva/negativa entre os três grupos e de grupos opostos, já que Orgulho/Vergonha e Orgulho/Humildade são pares diferentes, mesmo que a palavra “orgulho” seja a mesma nos dois pares.

Os capítulos 4, 5 e 6 estão diretamente envolvidos nesta terceira fase de desenvolvimento do Emocionário. O capítulo 4 elenca desafios encontrados na análise do corpus à luz da gramática do português e das palavras de emoção já contidas no Emocionário. Nele, discuto obstáculos e a dificuldade de anotar palavras que fazem referência a emoção. No capítulo 5, ao longo das explorações em corpus a partir de padrões, como o “sentimento de N”, descrevo a adição dos grupos Outras e Ausência, criados para dar conta de palavras que não se encaixavam em nenhum dos grupos já existentes. Também discuto possibilidades de trabalhos futuros que necessitariam de aprofundamento para adição de outros grupos com base nos campos lexicais de espiritualidade, família e nacionalismo. No capítulo 6, descrevo a metodologia desenvolvida para revisar os grupos Desespero, Esperança, Humildade, Pena, Satisfação, Saudade e Surpresa.

2.4

Humanidades Digitais e leitura distante

A área de Humanidades Digitais é concebida a partir da parceria entre as ciências humanas e recursos digitais para a criação de novas interpretações e potencialidades. É uma área de intensa interdisciplinaridade que utiliza majoritariamente dados abertos na condução de pesquisas.

Leitura Distante, ou *distant reading*, como originalmente foi denominada, é uma ferramenta da área de Humanidades Digitais idealizada por Franco Moretti (2008) para estudar literatura a partir de uma outra perspectiva. Em seu livro “A literatura vista de longe”, o autor explica sua proposta:

[...] aqui se fala de literatura: o objeto permanece mais ou menos aquele de sempre, diferentemente da recente virada do *new historicism* e, depois, dos *cultural studies* em direção a outros âmbitos de discurso. Mas a literatura é, não obstante, “vista de longe” no sentido de que o método de estudo aqui proposto substitui a leitura de perto do texto (a *close reading* da tradição de língua inglesa) pela reflexão sobre aqueles objetos artificiais (...). Objetos diferentes, mas todos resultado de um processo de deliberada *redução* e *abstração*. Em suma, de um distanciamento em relação ao texto em sua concretude. (Moretti, 2008, p. 7)

Portanto, a Leitura Distante consiste em distanciar-se analiticamente das obras literárias na busca por diferentes padrões, relações e formas; seja dentro de uma única obra, entre várias obras de um mesmo autor, ou mesmo em diferentes obras de diferentes autores. Nas palavras de Moretti,

Distant reading, chamei [...] este modo de trabalhar em que a distância não é um obstáculo, mas sim *uma forma específica de conhecimento*. A distância faz com que se vejam menos os detalhes, mas faz com que se observem melhor as relações, os *patterns*, as formas. (Ibid., p. 8)

Essa abordagem não deve ser encarada por pesquisadores dos estudos literários como uma ameaça ou um obstáculo, mas sim como uma nova forma de conhecimento. A análise a partir da Leitura Distante permite que o pesquisador dê um passo atrás para, metaforicamente, ser capaz de observar a floresta como um todo, distanciando-se da análise individual de cada árvore e ganhando uma perspectiva panorâmica sobre seu objeto de pesquisa. Santos et al. (2018) elencam dados que a Leitura Distante pode nos fornecer, como influências literárias entre autores ou escolas literárias, temas que estavam em foco em períodos específicos (e também quando saíram de foco), descrição de personagens e ligação entre elas, emoções presentes nos textos e um mapeamento de quem as exprime, entre outras possibilidades. A seção 2.5, a seguir, descreve algumas explorações envolvendo as áreas de estudos literários, corpus, HDs e a Leitura Distante, e a análise de menção de emoção.

2.5

Aplicação prática da leitura distante ao estudo de emoções em corpus

A pesquisa de Klinger et al (2016) intitulada *Automatic Emotion Detection for Quantitative Literary Studies: A case based on Franz Kafka's "Das schloss" and "Amerika"* visa a detectar diferentes emoções em dois romances de Franz Kafka. Para isso, os pesquisadores utilizam um dicionário de alemão criado

especificamente para este estudo: trezentas palavras foram atribuídas a um dos tipos de emoções fundamentais universais idealizadas por Robert Plutchik (“raiva”, “desgosto”, “medo”, “felicidade”, “tristeza”, “surpresa” e “satisfação”). A partir de então, foi possível detectar no corpus mudanças como aumento ou diminuição de sentimentos ao longo dos romances e quais sentimentos são mais presentes em cada personagem em diferentes momentos das obras, à luz do léxico de emoções compilado pelos próprios autores.

Grayson et al. (2016) pretendiam examinar as redes de interação dos personagens em nove romances de Jane Austen e Charles Dickens do século XIX. Para que a anotação de cada obra fosse realizada, foi criado um dicionário de personagem contendo uma única entrada para cada personagem, que continha seus possíveis apelidos ao longo da respectiva obra. Todas as coocorrências de cada obra foram contabilizadas, não só conversas diretas, de forma a capturar um maior número de interações e associações entre personagens. Um dos resultados foi observar que as sociedades construídas por Austen parecem ser mais compactas do que as de Dickens. Outro ponto foi o uso de muitas micronarrativas por Dickens, apesar de os autores acreditarem que a inserção de micronarrativas não afeta o enredo central, resumindo-se apenas a um estilo pessoal do autor.

A pesquisa de Heuser et al. (2016) analisa romances que se passam em Londres durante o século XIX para mapear as emoções da cidade. Os pesquisadores empregaram um programa de NER (Reconhecimento de Entidades Mencionadas) para selecionar as passagens que se referiam à cidade, como ruas ou bairros. Dentre as emoções iniciais, houve alta discordância, tanto entre os anotadores automáticos como entre os anotadores humanos, que serviam como grupo de controle. Optaram, então, por reduzir as opções de emoções para apenas duas, consideradas opostas: “medo” e “felicidade”. Os bairros mais frequentes nas obras foram Westminster e The City, o que, segundo os autores da pesquisa, refletia uma parcialidade nas representações da cidade de Londres. No entanto, ao apresentarem a “geografia do medo ou da felicidade” na cidade de Londres, eles descobriram que a maior parte das passagens foi anotada semanticamente como “emocionalmente neutra”. Mesmo se levadas em conta somente passagens marcadas com emoções, novamente West End e The City foram as regiões que mais apareceram marcadas com felicidade e medo, respectivamente.

Como visto nos resumos acima, as pesquisas realizadas oferecem múltiplas possibilidades de trabalho com o método de leitura distante. Se textos

são produzidos por pessoas, carregam a capacidade de falar sobre os valores das sociedades nas quais foram criados ou sobre as quais falam. Ao alternar as leituras distante e aproximada, o pesquisador terá a oportunidade de analisar achados em larga escala e, a partir de seu olhar especializado, detectar padrões e informações que talvez não saltariam dos dados em uma análise feita por uma leitura linear. Assim, é possível combinar seu olhar experiente com técnicas de áreas originalmente desassociadas aos estudos linguísticos para acessar um campo de análise mais vasto e ainda pouco explorado.

Portanto, a sinergia bem-sucedida entre os estudos com corpus e a Leitura Distante nas Humanidades Digitais, a análise de emoção e dados da literatura – ilustrada pelas pesquisas descritas nesta seção – revela que, quando análises qualitativas e quantitativas são combinadas, os pesquisadores têm muito a desfrutar das potencialidades dessas associações.

3

O Emocionário e Léxicos de sentimento em português: Análise de menção de emoção é Análise de Sentimento?

Este capítulo detalha a tarefa de Análise de Sentimento, suas possíveis aplicações comerciais e acadêmicas, e a contrasta com a categorização de sentimentos, escopo desta tese. Na seção 3.2, são enumerados e descritos alguns dos recursos usados para a tarefa de Análise de Sentimento na língua portuguesa: *WordNetAffectBR*, *OPLexicon*, *SentiLex*, *Multilingualsentiment* e *LIWC*. Aqui, são indicados apenas léxicos porque ao longo da revisão bibliográfica não foram encontrados projetos de anotação de palavras de emoção em português. Léxicos de sentimento atualmente podem ser desenvolvidos de duas formas: léxicos de opinião, ou de polaridades, compostos por palavras de diversos campos lexicais cuja orientação semântica é marcada como positiva, negativa ou neutra, à luz da Análise de Sentimento; ou léxicos de palavras de emoção, cujo principal embasamento teórico é o conceito de universalidade das emoções (ver seção 1.1 desta tese). As três abordagens mais comuns na construção de léxicos para detecção de orientação semântica são as baseadas em corpus, em dicionário e em tradução, como está demonstrado ao longo da descrição de cada léxico. Na seção 3.3, os léxicos de sentimento e/ou opinião são comparados ao Emocionário, cujo léxico é utilizado como ponto de partida desta pesquisa, para esclarecer as diferenças entre a Análise de Sentimento e a Análise de menção de emoção e responder à pergunta posta no título deste capítulo.

3.1

Análise de Sentimento

Sentiment Analysis, também conhecida como *Opinion Mining*, refere-se à tarefa nomeada em português como Análise de Sentimento (mais amplamente usada que o termo Mineração de Opinião) e consiste na detecção de orientação semântica (ou polaridade) de frases ou textos a partir da extração de informação

de um corpus. A orientação semântica, também conhecida como polaridade, pode ser positiva, negativa ou neutra. Há duas abordagens possíveis para a Análise de Sentimento: a abordagem baseada em aprendizado de máquina e a baseada em léxico. A Figura 2 é uma reprodução do diagrama de Taboada (2016:329) que resume os processos de cada abordagem.

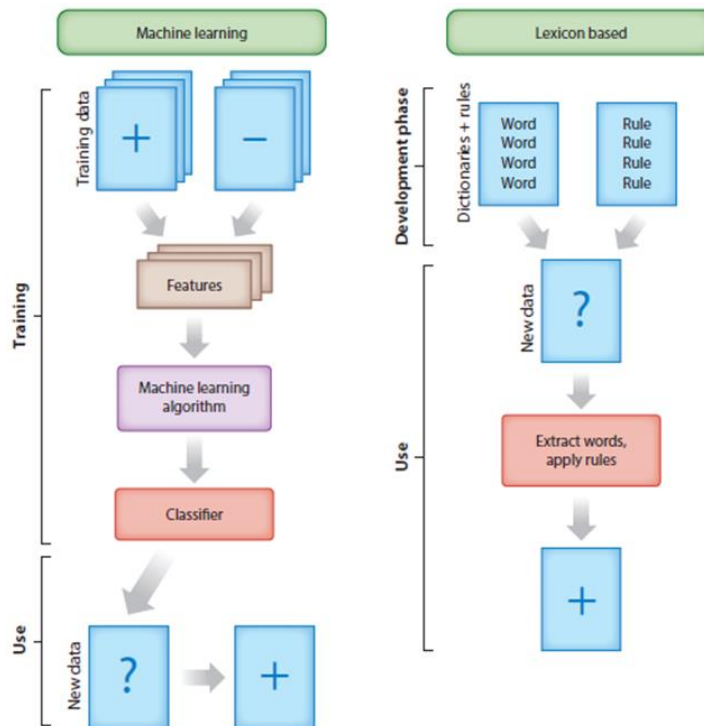


Figura 2 - Abordagens de aprendizado de máquina e baseada em léxico para análise de sentimento. (Taboada, 2016:329)

A abordagem baseada em aprendizado de máquina consiste na confecção de um classificador, a partir de um algoritmo, que é alimentado com um material de treino. Nesse classificador são inseridos os dados que o pesquisador deseja extrair de um *dataset* de forma automática. No caso da análise de sentimento, são anotadas as orientações semânticas de palavras. A principal vantagem desse método é a rapidez. No entanto, o classificador só pode ser desenvolvido em um material já anotado e, portanto, deve ser testado em materiais com contextos específicos para maior chance de atingir bons resultados. Assim, será necessário novo material de treino para cada contexto, que dependerá de nova anotação manual, diminuindo consideravelmente a rapidez do método.

Na abordagem baseada em léxico há maior equilíbrio na colaboração entre abordagens linguísticas e computacionais. A etapa de desenvolvimento do material envolve também a anotação linguística, que é usada como o *dataset* modelo para extração de informação em novos materiais e aplicação das

mesmas regras de anotação criadas para o *dataset* modelo. O produto é um léxico de opinião, que indica polaridades das palavras nele contidas. As implicações dessa abordagem envolvem como determinar a relevância de palavras, expressões e frases e como extrair padrões do novo material de análise. É também fundamental que o interesse por trás da construção desse léxico esteja bem definido: se sua aplicação terá propósitos comerciais ou um foco mais acadêmico.

A Análise de Sentimento pode ser aplicada nas mais diversas áreas com interesse comercial, que vai além de buscar entender como as pessoas sentem, mas também o que sentem. A análise de críticas e resenhas de livros, filmes, produtos e serviços de um modo geral, faz a Análise de Sentimentos diretamente em mecanismos de busca a fim de mensurar relevância para outros consumidores, mas também para vendedores ou produtores. A área também é muito utilizada na política, na exploração de tweets sobre partidos políticos, candidatos e personagens do cenário político; assim como para os indicativos de resultados de eleição e mais recentemente para delinear características dos autores das publicações on-line e suas interações. Para o mercado financeiro, a Análise de Sentimento pode auxiliar no acompanhamento da evolução de tendências e sentimentos de investidores em fóruns, por exemplo.

Já em uma abordagem de Análise de Sentimento mais mesclada entre interesses comerciais e acadêmicos, no campo da psicologia, Pestian et al. (2012) analisaram mensagens com teor suicida para tentar detectar quais casos poderiam ser reincidentes a partir de palavras de sentimento como *abuse*, *anger*, *sorrow*, *forgiveness*, *love* e *pride*. O pesquisador Junaid Nabi⁶, em um artigo para o Nexo Jornal em 2019, chamou atenção para a potencialidade do maior alcance das pesquisas e tratamentos para saúde mental quando áreas diferentes dialogam de forma colaborativa (neste caso, especialistas de IA, psicologia e, por que não, Análise de Sentimento?).

A Análise de Sentimento é uma área cuja maior ênfase tem sido na língua inglesa. Para Taboada (2016:339), trabalhos na área podem ser desenvolvidos para outras línguas de duas formas: a criação de métodos na própria língua a ser estudada, seja a partir da abordagem baseada em léxico ou aprendizado de máquina, ou a opção pela tradução de métodos já desenvolvidos para o inglês, seja por tradução de léxicos para a língua desejada ou pela tradução do *dataset*

⁶ Disponível em: <https://www.nexojornal.com.br/externo/2019/05/25/Uma-abordagem-inteligente-para-a-sa%C3%BAde-mental> Último acesso em março de 2022.

a ser analisado para o inglês para que a análise seja conduzida através de um sistema de análise de sentimento para inglês.

Na língua portuguesa, há trabalhos desenvolvidos a partir de traduções de corpora e/ou de léxicos de sentimentos, como o de Santos et al. (2014) que propõem uma Mineração de Emoções em textos multilíngues; e as pesquisas que resultaram nos léxicos descritos na seção 3.2, a seguir. No caso do Emocionário (apresentado na seção 2.3), o projeto de estudo e anotação das palavras de emoção possibilitou o desenvolvimento de um léxico de palavras de emoção partindo do zero. Na consulta ao AC/DC, o objetivo é observar como os sentimentos são descritos/materializados pelos próprios falantes de português, evitando a baixa concordância entre anotadores e teorias que argumentam a favor da universalidade dos sentimentos.

3.2

Léxicos de sentimento em língua portuguesa

3.2.1

WordNetAffectBR

O WordNetAffectBR⁷ (Pasqualotti e Vieira, 2008) é um léxico que indica polaridades de palavras de emoção em língua portuguesa. É descrito como uma base afetiva para a língua portuguesa, “construída para aplicações que utilizam vocabulário de palavras de emoções” (Pasqualotti e Vieira, 2008). Os autores afirmam que há dois ramos dentro da Computação Afetiva, uma das bases teóricas em que se insere o trabalho: emoções na interação homem-computador e síntese das emoções. Para elaborar o WordNetAffectBR, o foco recai sobre o primeiro campo, o das emoções na interação homem-computador, que “estuda mecanismos para reconhecer e expressar emoções por máquinas na interação homem-computador” (Pasqualotti e Vieira, 2008).

Foi construído a partir da combinação das ontologias lexicais WordNet, base Affect e WordNet Affect, originalmente em língua inglesa. A WordNet é uma ontologia lexical de língua inglesa que conta com mais de 200.000 palavras, de diferentes classes gramaticais, e synsets, que são conjuntos de palavras com um mesmo sentido. A base Affect é composta pelas mesmas classes

⁷ <https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/wordnet affectbr/>

gramaticais do WordNet, informações de ordem afetiva, baseadas no modelo Ortony et al (1988), nas teorias de emoções básicas de Elliot (1992), e teorias dimensionais que representam as valências afetivas positivas e negativas; e lexical e semântica, parcela que conta com um glossário, classe gramatical correspondente, relações de sinonímia e antonímia e correlações em inglês e italiano. A WordNet Affect comporta a parcela da WordNet relacionada a estados afetivos a partir das informações extraídas da base Affect e foi construída em duas etapas – a primeira foi a realização da correspondência dos termos e synsets entre as bases WordNet e Affect, e a segunda consistiu na correlação dos synsets da Wordnet já com as relações semânticas da base Affect adicionadas.

Bem como a versão em língua inglesa, a WordNetAffectBR apoia-se metodologicamente no modelo OCC (Ortony et al, 1988) e conta com traduções para o português das palavras e dos synsets já existentes na versão original. Ela foi elaborada respeitando a relação entre dados da base Wordnet Affect, da Wordnet e do estado afetivo das emoções anotadas como “Emo”. É interessante ressaltar que os autores do artigo expõem que diversas palavras em inglês foram traduzidas por expressões e, por isso, foram retiradas do material em português, já que seu interesse era somente em palavras. A título de exemplo, cito a palavra “*gloating*”, que foi traduzida como “prazer pelo sofrimento alheio” e “alegria pelo sofrimento alheio”. Ou seja, o sentido afetivo foi mantido, mas a palavra em inglês foi traduzida para uma expressão em português, por ausência de uma palavra equivalente em sentido na tradução. Ainda que o processo de tradução tenha sido conduzido por tradutoras profissionais e experientes em trabalhos desse tipo, como consta no artigo, no processo de tradução para listas de palavras perde-se muito do que tanto a língua fonte como a língua alvo têm a oferecer, porque muitas palavras não possuem palavras equivalentes em uma ou outra língua.

O léxico apresentou 457 palavras depois da fase final da tradução e, após o processamento e retiradas as repetições, o resultado obtido foi um léxico de 289 palavras de emoção com polaridades designadas entre positivas (“admiração”, “carinhoso”, “orgulho”) ou negativas (“inveja”, “pavor”, “vergonha”), que está disponível on-line.

3.2.2 OPLexicon

O OpLexicon⁸ (Souza et al., 2011) é um léxico de opinião idealizado para aplicação na área de análise de sentimento. Ele foi elaborado a partir da consideração de que as três abordagens mais comuns na construção de léxicos para detecção de orientação semântica são as baseadas em corpus, em dicionário e em tradução. Logo, os pesquisadores elaboraram dois *seed sets* para serem aplicados em cada abordagem e tomaram por inspiração os trabalhos de Turney (2002), Kamps et al. (2004) e Mihalcea et al. (2007) (apud Souza et al., 2011), além de ferramentas linguísticas para compilar palavras ou termos que carregassem opinião. Dessa forma, construíram três léxicos de opinião que foram combinados para criar um léxico único para a língua portuguesa, aproveitando as vantagens de cada abordagem e sobrepondo os obstáculos que cada uma oferece com auxílio das outras.

Para o léxico baseado em corpus, apesar de terem se inspirado no trabalho de Turney (2002) que parte de um corpus criado por um mecanismo de busca na internet, os pesquisadores optaram por usar um corpus já processado e documentado. Esse corpus conta com 346 críticas de filmes publicadas nos sites CinePlayers e Cinema com Rapadura e escritas em português do Brasil e 970 textos de jornal extraídos do corpus PLN-Br CATEG, totalizando aproximadamente um milhão de palavras. O corpus foi anotado com o método *Pointwise Mutual Information* e foram selecionadas apenas expressões com polaridades acima da polaridade média dos conjuntos de palavras disparadoras para obter resultados mais precisos.

O léxico baseado em dicionário tomou como base o método de Kamps et al. (2004), que para calcular as distâncias entre palavras com antônimos, como por exemplo “bom” e “ruim”, e os conjuntos de palavras disparadoras, e assim selecionaram as palavras cujas distâncias eram menores ou iguais a três. O tesouro *TeP* (Dias da Silva et al., 2001) , que conta com 44.077 palavras anotadas com sinonímia e antonímias, foi usado como recurso lexical.

Para a abordagem baseada em tradução do terceiro léxico, os autores utilizaram o recurso *English Opinion Lexicon* (Hu e Liu, 2004), com aproximadamente 6.800 entradas. Como solução para simplificar e acelerar o processo de construção do léxico e reduzir a carga de trabalho manual, a avaliação foi feita entre inglês e português e as palavras e expressões foram

⁸ <https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/oplexicon/>

traduzidas de forma automática com o *Google Translate Online*. Nesse processo, todas as traduções foram aproveitadas, tendo sido descartadas manualmente expressões que não foram traduzidas (por erros de digitação no texto original, por exemplo).

A construção dos léxicos para as três abordagens resultou em uma lista de 359 expressões para o léxico baseado em corpus; 2.400 palavras para o léxico baseado em dicionário e 4.909 expressões para o léxico baseado em tradução. A combinação dos três léxicos gerou o OpLexicon, com 7.077 lemas com anotação de orientação semântica (positiva ou negativa). Os casos de discordância entre polaridades de cada um dos três léxicos foram sempre resolvidos descartando a polaridade gerada pela fonte do léxico traduzido ou medindo a confiabilidade das fontes. Visto que o léxico baseado em tradução é proveniente de recursos para outras línguas, ele foi considerado o menos confiável.

3.2.3

LIWC / Brazilian Portuguese LIWC 2007 Dictionary

O LIWC⁹ (Pennebaker et al., 2001) é um software para análise de textos em inglês baseado em um léxico que classifica as palavras de acordo com categorias “psicologicamente significativas”. A versão mais abrangente do léxico é de 2007 (Pennebaker et al., 2007) e contém em torno de 127.149 palavras.

Ele é organizado em quatro categorias gerais de descritores (contagem de palavras, palavras por oração, porcentagem de palavras reconhecidas pelo dicionário e porcentagem de palavras com mais de seis letras); 22 categorias de dimensões linguísticas, contabilizando pronomes, artigos, verbos auxiliares, etc no texto); sete categorias de interesses pessoais, como trabalho, lar, atividades para lazer; três dimensões paralinguísticas, como indicadores de desvios de fluidez no texto; doze categorias de pontuação; e 32 categorias compreendendo construtos psicológicos, dentre elas a categoria de “*Affective Processes*” (*happy, cried, abandon*), que por sua vez compreende as subcategorias “*Positive emotion*” (*love, nice sweet*), “*Negative emotion*” (*hurt, ugly, nasty*), “*anxiety*” (*worried, fearful, nervous*), “*anger*” (*hate, kill, annoyed*) e “*sadness*” (*crying, grief, sad*). A categoria “*Affective Processes*”, e as subcategorias correspondentes

⁹ http://143.107.183.175:21380/portlex/images/arquivos/liwc/LIWC2007_Portugues_win.dic.txt

somam 2.196 lemas do LIWC, que se repetem nas contas das categorias. A forma “*cried*”, por exemplo, está incluída nas categorias “*sadness*”, “*negative emotion*”, “*overall affect*”, “*verb*” e “*past tense verb*” (Pennebaker et al, 2007:4).

Para a classificação, inicialmente seriam identificadas palavras cujos sentidos se aproximassem de dimensões emocionais e cognitivas básicas e estudadas nos campos da psicologia. A categorização das palavras acabou acontecendo em quatro etapas, denominadas “*Word Collection*”, “*Judges’ Rating Phases*”, “*Psychometric Evaluation*”, e “*Updates and Expansions*” (Pennebaker et al., 2007:7-8). A primeira etapa consistiu na elaboração das listas de palavras para cada categoria, que foram retiradas de diferentes fontes, como dicionários comuns, tesouros e dicionários especializados, como por exemplo em emoção ou processos afetivos no caso da categoria “*Affective Processes*”. A partir das listas iniciais, foram realizadas sessões de *brainstorming* com peritos para adicionar palavras relevantes a uma ou mais categorias.

A segunda etapa consistiu na avaliação de todo o conteúdo do LIWC por três juízes – primeiro foram revisadas quais palavras do dicionário seriam mantidas ou excluídas nas categorias “*Psychological Processes*”, “*Personal Concerns*” e “*Relativity*” (exceto a subcategoria “*verb tense*”). Em seguida, exceto para a categoria “*Standard Language Dimensions*” (que inclui artigos, pronomes e preposições), foi feita a avaliação para a permanência, adição ou exclusão de palavras das demais categorias e subcategorias. Para que isso ocorresse, os juízes receberiam listas de palavras organizadas em ordem alfabética das categorias e deveria haver um consenso de ao menos dois dos três juízes dessa etapa: primeiro para avaliar a inserção, manutenção ou exclusão da palavra na categoria, e depois em alguma das subcategorias. Ao final, na categoria *Cognitive Process* e respectivas subcategorias, por exemplo, a concordância entre os juízes variou entre 93% e 100%.

A terceira etapa ocorreu ao longo da década de noventa com o objetivo de limpar o software e os dicionários. Após a análise de arquivos com mais de oito milhões de palavras, em diferentes fases e por diferentes grupos de juízes, e com o auxílio da ferramenta WordSmith, categorias, subcategorias e palavras com baixa frequência de uso foram omitidas, como a de “*Communication Verbs*” ou “*Metaphysical*” e outras categorias, como as de interesse pessoal e dimensões relativas, foram adicionadas ao LIWC.

Finalmente, a quarta etapa envolve as atualizações e expansões desde a versão inicial do LIWC e tem como produto a criação de novas categorias e adição de novas palavras às listas já existentes a partir de uma análise das 2000

palavras de uso mais frequente do léxico. Um exemplo é a adição de categorias para conjunções, advérbios, quantificadores e verbos auxiliares na última versão, motivada por estudos que ratificam a importância de palavras gramaticais em análise de processos psicológicos.

A versão em português foi construída por três equipes que fizeram traduções manualmente para construir o *Brazilian Portuguese LIWC 2007 Dictionary*¹⁰. Foram inseridas automaticamente conjugações, utilizando um dicionário, e as categorias e subcategorias preexistentes do LIWC, e as traduções manuais não foram revisadas. A presença da dimensão afetiva faz do LIWC um recurso interessante para a Análise de Sentimento, e nesse contexto foi feita uma avaliação do LIWC desenvolvido para a língua portuguesa. Para a avaliação da versão em português do LIWC (Balage et al, 2013), foram consideradas apenas as categorias que correspondem diretamente a polaridades: “*Positive emotion*” e “*Negative emotion*”, já que foi comparada ao OPLexicon e ao SentiLex-PT, que são léxicos de polaridade e contêm apenas categorias de orientação semântica. O resultado da avaliação foi que o LIWC em português brasileiro ficou na média dos outros léxicos e tem melhor desempenho detectando emoções positivas que negativas.

3.2.4 SentiLex-PT

O SentiLex¹¹ (Carvalho e Silva, 2015), assim como o OPLexicon, é um léxico de opinião que foi construído a partir da combinação de diferentes abordagens de construção de léxico. Neste caso, foram usadas as abordagens baseadas em corpus, com extração semiautomática de possíveis adjetivos humanos, e em aprendizagem automática para filtrar quais adjetivos seriam, de fato, humanos. A classificação de polaridade foi feita principalmente de acordo com a distância das palavras com relações de sinonímia em um gráfico, tomando como base dicionários de sinônimos em língua portuguesa.

Além de ter sido idealizado para o português de Portugal, sua característica mais expressiva e que mais o diferencia do OpLexicon, é que ele tem como foco predicadores humanos. Os 7.014 lemas que compõem o léxico pertencem a diferentes classes de palavras, como adjetivos, nomes, verbos ou

¹⁰ <http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>

¹¹ http://dmir.inesc-id.pt/project/SentiLex-PT_02

locações desde que possam modificar substantivos humanos que desempenhem função de sujeito ou de complemento. Ou seja, sendo composto por palavras variadas que se referem a pessoas, com a indicação da polaridade atribuída tanto à pessoa que predica quanto ao alvo da predicação. Os autores ilustram a característica com o caso do adjetivo “frágil”, que pode modificar nomes concretos, abstratos e de natureza humana (respectivamente *cobertura frágil*, *posição frágil* e *indivíduo frágil*), sempre com orientação semântica negativa. Nesse caso, apenas «frágil» como modificador de indivíduo faz parte do SentiLex-PT e «há, portanto, termos de sentimento comuns na língua que, por serem meros modificadores de nomes não humanos, não estão contemplados neste léxico (e.g. *nítido*); pelo contrário, outros predicadores, como os ilustrados anteriormente, encontram-se atestados, apesar da sua ambiguidade inerente” (Carvalho e Silva, 2015:428). A classe gramatical com o maior número de palavras é a de adjetivos, com a maioria dos predicadores (67%) sendo de polaridade negativa.

As entradas dos lemas no SentiLex-PT contêm informações sobre natureza sintática do predicador (transitivo ou intransitivo); natureza semântica dos argumentos (categoria humana, neste contexto do SentiLex-PT); polaridade do predicador quanto ao alvo que ele modifica; como foi atribuída a polaridade (manual ou automaticamente); e informações de lema formas flexionadas correspondentes.

O SentiLex-PT comporta dois dicionários, um descrevendo lemas e um para as formas flexionadas. Cada lema comporta informações relacionadas a categoria gramatical (adjetivo, substantivo, verbo e expressão idiomática) e atributo de sentimento, compreendendo a polaridade (positiva, negativa ou neutra), o alvo da polaridade (sendo ele um substantivo humano usado como predicativo do sujeito ou aposto nominal), e a classificação da polaridade (se a polaridade do lema foi atribuída de forma manual ou automática).

3.2.5 Multilingualsentiment

A partir da lacuna de léxicos de sentimento disponíveis publicamente na maioria das línguas encontrada por Chen e Skiena, o Multilingualsentiment¹²

¹² <https://sites.google.com/site/datascienceslab/projects/multilingualsentiment>

(Chen e Skiena, 2014) se propõe a elaborar um léxico internacional de sentimentos que comporta mais de oitenta línguas.

O processo de criação do léxico se deu em duas etapas principais e teve como característica de criação mais marcante a utilização de recursos já existentes que, combinados, formaram o Multilingualsentiment. A primeira etapa partiu do “*The Polyglot project*” idealizado por Al-Rfou et al. (2013, apud Chen e Skiena, 2014:384), que identificava as 100 mil palavras mais comuns das páginas da Wikipedia em cada língua, independentemente de serem do campo lexical de emoções. Foram coletadas 7.741.544 palavras de 136 línguas. A fim de estabelecer links semânticos, essas palavras foram usadas como núcleos de uma lista base em formato de gráfico.

Os seguintes recursos deram início à segunda etapa do projeto: *Wiktionary*, *Google translation API*, *transliteration links* e *WordNet*. Após a integração, os autores conectaram seus dados às palavras do gráfico construído na primeira etapa do projeto. A *Wiktionary* gerou aproximadamente 20% do total de links com o gráfico; o *Google translation API* forneceu pouco mais de 50% dos links, estabelecendo conexões entre 3,5 milhões dos vértices do gráfico; os *transliteration links*, que relacionaram palavras de línguas que compartilham escritas iguais ou similares, formaram 22,1% dos links; e a *WordNet* contribuiu com 5% dos links, a partir de 100 mil pares de sinônimos e antônimos na língua inglesa.

As polaridades das palavras do Multilingualsentiment começaram a ser atribuídas a partir dos léxicos em língua inglesa, por meio de um algoritmo de anotação e um de propagação de grafo, que se mostrou mais eficiente por escolher sempre o melhor par de palavras entre os vértices. Assim, os links semânticos se criavam a partir da extensão da polaridade a palavras adjacentes em significado. Primeiramente, apenas antônimos criavam links com polaridade negativa – no caso de um vértice que gerasse links positivos e negativos, a polaridade era considerada neutra. No geral, a polaridade de vértices que recebiam links múltiplos era calculada a partir de um balanço entre as polaridades dos links. Ao final, foram conduzidos testes em um formato de *grid search* para confirmar a precisão das polaridades de léxicos que não eram em língua inglesa.

A Tabela 1, a seguir, destaca as principais características dos léxicos de sentimento descritos neste capítulo, seus tamanhos e o endereço virtual onde cada um pode ser encontrado.

Léxico:	Principais características:	Tamanho:	Disponível publicamente:
WordNetAffectBR (Pasqualotti e Vieira, 2008)	- Construído para o português brasileiro a partir de traduções da base em língua inglesa.	289 palavras	Sim ¹³
OPLexicon (Souza et al., 2011)	- Combinação de três léxicos construídos a partir de diferentes abordagens.	31.668 palavras e/ou expressões.	Sim ¹⁴
LIWC (Tausczik e Pennebaker, 2010; Balage et al., 2013)	- A versão em português só compreende as categorias que contêm polaridades das palavras do léxico: “ <i>Positive emotion</i> ” e “ <i>Negative emotion</i> ”.	127.149 palavras, sendo 28.592 as que compõem a categoria “ <i>affective processes</i> ”	Sim ¹⁵
SentiLex-PT (Carvalho e Silva, 2015)	- Idelizado para o português de Portugal; - Foco em predicadores humanos.	7.014 lemas	Sim ¹⁶
Multilingualsentiment (Chen e Skiena, 2014)	- A construção partiu de diferentes recursos já existentes como <i>Wiktionary</i> , <i>Google translation API</i> , <i>transliteration links</i> e <i>WordNet</i> .	7.741.544 palavras (3.953 das quais em português) em 136 línguas diferentes	Sim ¹⁷

Tabela 1 - Léxicos de sentimento e suas principais características.

Inicialmente, a tabela também continha uma coluna para ser preenchida com o que é considerado emoção em cada léxico. No entanto, após as análises individuais, o WordNetAffectBR foi o único léxico que relatou explicitamente que o Modelo OCC (Ortony et al, 1988) foi utilizado como parâmetro de emoção. Os demais trabalhos optam pelo caminho de detectar “emoções positivas” ou “emoções negativas”, sem adentrar pela seara de “emoções básicas”, ou modelos já validados na literatura sobre emoções (descritos na seção 1.1 desta pesquisa) que tentam identificar os tipos de emoção existentes.

¹³<https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/wordnet affectbr/>. Último acesso em outubro de 2019.

¹⁴<https://www.inf.pucrs.br/linatural/wordpress/recursos-e-ferramentas/oplexicon/>. Último acesso em outubro de 2019.

¹⁵http://143.107.183.175:21380/portlex/images/arquivos/liwc/LIWC2007_Portugues_win.dic.txt

¹⁶http://dmir.inesc-id.pt/project/SentiLex-PT_02 - versão 1, download feito em 23/05/2019

¹⁷<https://sites.google.com/site/datasciencesslab/projects/multilingualsentiment>

3.3

Por que não é possível comparar os léxicos ao Emocionário

Em uma tentativa de comparação entre os léxicos em nível quantitativo, a fim de registrar a incidência de palavras iguais entre o Emocionário e os demais léxicos, a Tabela 2 mostra a proporção de ocorrências equivalentes aos lemas do Emocionário em cada um dos outros léxicos:

Léxico:	Equivalências / Total de palavras ou formas do léxico:	Exemplos em comum com o Emocionário:	Exemplos que não estão no Emocionário:	Exemplos que aparecem somente no Emocionário:
WordNetAffectBR (Pasqualotti e Vieira, 2008)	133 / 289	Infeliz; lamentável; raiva; solidariedade; veneração.	Afetuosos; beatífico; estressado; presunçoso; vermelho.	Coragem; desgosto; possesso; predileto; respeitoso.
OPLexicon (Souza et al., 2011)	2.626 / 31.668	Cabisbaixo; destemido; envaidecer; lenir; querido.	Ermo; ferino; heterogêneo; maligno; rasurar.	Cobiça; consternação; desopilar; expectativa; raiva.
LIWC (Pennebaker et al., 2001, 2007; Balage et al., 2013)	4.915 / 28.592	Aflito; exultação; frustrado; nostalgia; prazer	Cerimônia; infração; noção; original; reduzir.	Alardear; almejar; desfrute; prazeroso; recompensar.
SentiLex-PT (Carvalho e Silva, 2015)	305 / 6.321	Acanhado; cobiçado; garrido; merencório; querido	Garoto; idoso; orelhudo; relapso; zeloso.	Encapelado; humilde; pesar; proceloso; sanhoso.
Multilingual-sentiment (Chen e Skiena, 2014)	315 / 3.953	Animado; fé; herói; orgulho; sensível.	Aprovado; elite; inteligência; magia; provar.	Compaixão; eufórico; ingratidão; inveja; ternura.

Tabela 2 - comparação entre os léxicos para o português e o Emocionário.

A comparação das palavras presentes em cada lista é de utilidade limitada, já que os números não conseguem aprofundar nenhuma interpretação além de quantidades de equivalências. Visto que se trata de listas com diferentes palavras agrupadas por diferentes abordagens e retiradas de diferentes fontes, analisar que 46% das palavras do WordNetAffectBR enquanto menos de 5% das palavras pertencentes ao SentiLex estão contidas no Emocionário, por exemplo, não é muito informativo, pois os objetivos de cada material são diferentes.

Por outro lado, a comparação também evidenciou como recursos construídos a partir de abordagens distintas são englobados por um mesmo

rótulo. Após analisar os léxicos de opinião construídos para a tarefa de Análise de Sentimento, foi possível perceber que a natureza das aplicações dos léxicos descritos neste capítulo é diferente da motivação da construção de um léxico de palavras de emoção como o Emocionário, já que seu principal interesse não é detectar polaridades de palavras, mas sim identificar e categorizar palavras de sentimento escolhidas por falantes de português.

A dificuldade de lidar com a expressão de emoções talvez explique a maior quantidade e popularidade do léxico de polaridades e/ou opiniões do que léxicos de palavras de emoções ou sentimentos propriamente ditas, e principalmente de projetos de anotação que tentem dar conta da menção de emoção nas línguas. No processo de construção de um léxico de palavras de emoção a partir da observação e exploração de corpora juntamente com o projeto de anotação das emoções na língua portuguesa, é possível ir além da detecção de orientação semântica passando a uma categorização de sentimentos.

Além disso, os léxicos de opinião/polaridades são carregados de apelo comercial em sua construção bem como nas possíveis aplicações – como pesquisas de avaliação sobre um produto ou serviço e outros usos mencionados na seção 3.1 –, enquanto a construção de um léxico de palavras de emoção que não é feita à luz da universalidade das emoções, como no caso do Emocionário, pode estar mais relacionada à descrição linguística e ao campo das Humanidades Digitais.

A tarefa de Análise de sentimento, portanto, se faz limitada dentro do objetivo desta tese. Desta forma, opto por considerar as tarefas atribuídas à presente pesquisa como sendo análise da menção de emoção e categorização de sentimentos.

4 Características e desafios da análise de emoção

Neste capítulo são detalhadas características e desafios da menção de emoção para demonstrar ao leitor a dificuldade da tarefa de analisar aspectos semânticos dentro da língua, com foco em descrição de emoção. As características, bem como os desafios, são expostos através de exemplos curiosos, desafiadores ou passíveis de discussão encontrados durante a análise, também como forma de dimensionar o problema que me proponho a tratar ao longo desta pesquisa. É importante destacar que os exemplos citados neste estudo foram integralmente encontrados no processo de análise do léxico de palavras de emoção do AC/DC¹⁸, onde é possível alternar o foco de busca entre leitura distante (apresentada na seção 2.4 desta tese) e leitura aproximada graças às ferramentas de busca que fazem parte do AC/DC (apresentadas na seção 2.2 desta tese). Fundamentalmente, três questionamentos direcionam a análise deste capítulo:

- (i) Qual o caminho utilizar para confirmar se uma palavra é de emoção ou não? (ver seções 4.2, 4.3, 4.5 e 4.7);
- (ii) Em que casos a palavra pode ser considerada uma descrição de emoção? (ver seções 4.1, 4.2, 4.3, 4.4, 4.7 e 4.8);
- (iii) Em que grupo (ou grupos) de emoção cada palavra se encaixa? (ver seções 4.6 e 4.7).

4.1 Expressões idiomáticas e expressões de várias palavras

A lexicalização é o “processo de opacificação que as palavras vão sofrendo ao longo de sua permanência e uso na língua” (Sandmann, 1997:67). A diferença entre um conjunto de palavras passar ou não por um processo de lexicalização está, portanto, no uso da expressão pelos falantes, o que pode aumentar ou diminuir o grau de opacidade das expressões. Por outro lado, a

¹⁸ Nas respectivas versões dos corpora nele contidos, entre outubro de 2019 e fevereiro de 2020.

pesquisa com corpus pode possibilitar a detecção e anotação dessas expressões a partir do conhecimento do pesquisador falante (neste caso) do português, que já possui conhecimento prévio para desconfiar de que uma expressão encontrada em uma busca no corpus trata de uma expressão lexicalizada. Nestes casos, a pesquisa com corpus facilita a busca e localização dessas expressões lexicalizadas, por tratar grandes conjuntos de texto de forma quantitativa e não dependendo de uma leitura linear, como já explicado na seção 2.4 desta tese.

A opacidade semântica é muito visível na exploração de descrição de emoção com expressões do corpo humano¹⁹, porque vemos que os falantes utilizam “cabeça fresca” ou “esfriar a cabeça” para descrever “alívio”, mas não “cabeça gelada” ou “cabeça refrescada” (Leitão de Almeida, 2009:203); assim como utilizam “com quatro pedras na mão” e “em pé de guerra” para expressar “fúria” ou “ódio”, ao invés de trocarem “pedras” por “rochas” ou “guerra” por “briga”. São expressões que se cristalizaram de modo que o sentido não está apenas na junção das palavras, mas sim na combinação entre forma e sentido. Vale (2013) também aponta que a utilização de expressões fixas possuem cargas de expressividade mais fortes, funcionando como uma espécie de “reforço da informação” (2013:266).

As expressões “Rua da amargura” e “fundo do poço”, por exemplo, carregam o sentido de “infelicidade” e já estão cristalizadas na língua portuguesa. São usadas pelos falantes com a conotação metafórica, como mostram os Exemplos 7 a 12,

Exemplo 7: par=FSP950104-090-994: Dominique obtém o divórcio «pela incapacidade do marido de consumir o casamento» e joga Karol na **rua da amargura** .

Exemplo 8: par=PUBLICO-19940310-007-111: E se o seu prestígio anda pelas **Ruas da amargura**, continua a ser um dos programas de maior audiência.

Exemplo 9: Talvez que extinto o mundo, apagados todos os astros, feito o universo treva, talvez ela ainda exista, e os seus soluços sinistramente ecoem na total ruína, rua das lágrimas, rua do desespero - interminável **rua da Amargura** .

Exemplo 10: par=t378 txt-6365: A situação já era gravíssima, mas a execução planejada de uma vereadora marca um novo **fundo do poço**.

Exemplo 11: " Ninguém pode se esquecer que foi o setor exportador que tirou o país do **fundo do poço** da recessão do Plano Collor e que 60 % da nossa pauta de exportação é composta de produtos industrializados. "

¹⁹ Uma exploração em corpora do AC/DC conectando a descrição de emoção e expressões lexicalizadas com o corpo humano a partir da anotação do projeto Esqueleto (Freitas et al., 2015) está descrita na seção 5.2 desta tese.

Exemplo 12: par=PUBLICO-19950621-128-1238: Seria muito duro se depois desta fractura, deste período em que sentimos que chegámos ao fundo do poço, viéssemos ao de cima para encontrar a mesma Argélia do passado.

Mas “fundo do poço” também aparece no corpus com sentido literal, como nos Exemplos 13 a 15.

Exemplo 13: par=ACO-Infelizes:_Histórias_vividas-28: Passava dias a olhar o **fundo do poço** onde a água se mostrava estagnada, negra, e ao mesmo tempo fascinante -- como a prometter-lhe descanço no interior da terra bem amada.

Exemplo 14: par=HC-O_monstro_e_outros_contos-1: Era preciso, porém, olhar para o **fundo do poço**, e receber com os olhos, de longe, o seu prometido; a ansiedade dominou-a, curvando-a sobre o abismo.

Exemplo 15: par=ext776350-soc-93b-2: Pedro Vieira, da associação ecologista Quercus, disse que a deposição de lamas -- que estão no centro das suspeitas de contaminação da água -- no **fundo do poço** de captação é lenta e, por isso, acha estranho que só agora os serviços tenham dado pela sua existência.

Esse jogo de sentidos é facilitado pelo fato de nenhuma das palavras ter o sentido de emoção separadamente. Esta seção destaca a importância das expressões lexicalizadas na análise semântica de emoção, bem como o contexto no qual elas estão inseridas, para que o projeto de anotação, a partir do estudo da língua, possa dar conta das expressões usadas para descrever emoção e enriqueça o léxico de emoção, diferentemente do que acontece na construção de um léxico composto apenas por palavras.

Além de acontecer em expressões lexicalizadas, como ilustraram “fundo do poço” e “rua da amargura”, a multiplicidade de sentidos também acontece em palavras únicas, mencionada na seção 4.2, a seguir.

4.2 Polissemia

Há dois tipos de polissemia relativas à menção de emoção reconhecidos na observação dos corpora do AC/DC: (i) palavras polissêmicas cujos sentidos variam entre palavra de emoção ou não e (ii) a polissemia de sentidos de emoção em uma mesma palavra.

4.2.1

Palavras polissêmicas cujos sentidos variam entre palavra de emoção ou não

A polissemia sempre foi um desafio para os estudos de lexicografia, e não é diferente nos estudos sobre expressão de emoção. Ao compararmos as palavras misericordioso e súbito, por exemplo, é perceptível que a primeira conta com menos maleabilidade semântica que a segunda, pois “súbito” pode ter outros sentidos além do sentido emocional “surpresa”, como “repentino” e “rápido”, ilustrados nos exemplos 16 a 18 abaixo:

Exemplo 16: par=59: Já repararam que as pessoas, quando se apaixonam, ganham uma **súbita** elasticidade ?

Exemplo 17: par=t507-4: Alguns dos sintomas de infecção de um smartphone são semelhantes aos de um computador tradicional, tais como a quebra **súbita** na performance do equipamento, o surgimento de aplicações estranhas instaladas no sistema ou a alteração de configurações no browser, pelo que o utilizador deverá também estar atento a estes sintomas .

Exemplo 18: par=fut-BR-GazetaEsp-2000-67202: A **súbita** lua-de-mel entre o atacante do Vasco e o técnico parou por aí.

O adjetivo “misericordioso” faz menção a emoção sem a necessidade de recorrermos ao contexto para confirmar, pois é menos polissêmico, enquanto a palavra “súbito”, como ilustraram os Exemplos 16 a 18, depende do contexto em que é utilizada para ser interpretada como uma palavra que descreve ou não emoção.

Essa dependência do contexto para descrever ou não emoção acontece também com os verbos “chocar” e “inchar”, ilustrados nos Exemplos 19 a 26, que podem respectivamente significar²⁰ “batida” ou “surpresa”, e “aumentar de tamanho” ou “orgulho”:

“chocar” com sentido de “batida”:

Exemplo 19: par=ext869911-soc-96a-2: «Foram bebendo um copo aqui e outro ali e, depois, **chocaram** de frente com uma camioneta», é a explicação plausível apresentada pelo presidente da junta .

Exemplo 20: par=FSP940201-122-864: De acordo com a versão do Dersa, o veículo teria se **chocado** primeiro com a placa de sinalização de altura que fica na marginal .

“chocar” com sentido de surpresa:

Exemplo 21: <p>: Analogamente, alguns dos crimes que mais **chocaram** o país.e talvez por isso mesmooocorreram nos mais altos círculos sociais .

Exemplo 22: <p>: Sem papas na língua, Marguerite **choca** a cada entrevista, confessa seu apego ao dinheiro e às jóias, tão duramente conquistados, revela as

²⁰ De acordo com as entradas dos dicionários virtuais Caldas Aulete e Priberam, último acesso em janeiro de 2020.

taras mais inconfessáveis suas e dos outros, sofre processos por calúnia e, quando os perde, paga mas não se cala .

“inchar” com sentido de “aumentar de tamanho”:

Exemplo 23: par=3320: <> Três horas seguidas, minhas pernas todas escalavradas das porradas que levei, o dedão do pé direito **inchado**, talvez quebrado. </>

Exemplo 24: par=2429: <> Se a mangueiras estivessem carregadas -- atiravam-se às mangas verdes ou simplesmente **inchadas**, não dando tempo a que elas amadurecessem. </>

“inchar” com sentido de “orgulho”:

Exemplo 25: Voltemos ao Euro do nosso contentamento que nos permite agora **inchar** o peito e proclamar aos quatro ventos e em todas as línguas faladas e escritas deste planeta e de todos os outros que estão espalhados pelo universo que temos uma moeda que vale mais que o próprio dólar americano .

Exemplo 26: <p>: Veja Vossa Senhoria até onde chega a grandeza de alma do Senhor Vilares em patrocinar aquele honrado Reitor quando melhor obraria este senhor se se lembrasse do que eu lhe fiz em outro tempo porém os tempos mudam e com os empregos **incham** os homes e depois de inchados cegam dos olhos para não ver o que lhe fica atrás .

Essa característica, por outro lado, ilustra as limitações de uma análise ou identificação de emoções em textos baseadas apenas em léxicos, que desconsideram o contexto.

4.2.2

Polissemia de sentidos de emoção em uma mesma palavra:

A multiplicidade de sentidos em uma mesma palavra de emoção pode significar que o lema é vago entre grupos de emoção diferentes ou que o sentido emocional daquela palavra se faz, na verdade, pela combinação de várias palavras de emoção. Essa característica nas palavras de emoção foi tratada em Santos e Mota (2015) em relação à polissemia de palavras que descrevem admiração, que pode representar a “admiração-respeito” e a “admiração-espanto” na língua portuguesa. O verbo “admirar”, por exemplo, quando usado de forma pronominal, significa “surpresa” na maior parte dos casos; em sua forma transitiva significa “veneração”; e seu uso apassivado pode significar os dois, dependendo dos verbos auxiliares associados. Em vários outros casos, porém, a estrutura sintática não é suficiente para dar conta dos possíveis sentidos, além de haver casos em que a palavra é caga em vários sentidos, como ilustrado nos exemplos 27 a 29 abaixo, retirados de Santos e Mota (2015:72)

Exemplo 27: Esse Hilário tinha na montra da oficina dele, em 1945, um carro feito por ele que quem passasse por lá admirava aquela arte.

Exemplo 28: aqueles dois recentes conhecimentos de Albernaz, embevecidos, boquiabertos e invejosos diante das proezas imaginárias daqueles três militares

Exemplo 29: Ângelo aos quinze anos já embasbacava os seus ingênuos professores

Em seu estudo, as autoras concluíram que o sentido de “espanto” é mais frequente que o de “respeito” tanto no âmbito de coocorrências como também nas opções de palavras possíveis de utilização no português. No entanto, reconhecem que há casos difíceis que compreenderiam uma junção dos dois sentidos, como nos exemplos 27 a 29.

A polissemia, independente de ocorrer entre sentidos emocionais ou não na palavra, é uma propriedade constitutiva na linguagem natural, e precisa ser levada em consideração na anotação do corpus. Destaco aqui a importância de o pesquisador mesclar a leitura distante com a leitura linear de linhas de concordância no corpus para que sejam esclarecidos sentidos polissêmicos, sempre validando os achados quantitativos a partir da confirmação com o contexto.

4.3

Classes gramaticais e a variação de sentido emocional

Algumas palavras homógrafas pertencentes a diferentes classes gramaticais podem possuir ou não sentido emocional ou mesmo ter variação de sentido emocional. Desse grupo de palavras, comento aqui os casos de “pesar”, “sentido”, “obrigado” e “querido”.

O substantivo “pesar” tem como sinônimos “tristeza”, “pena”, “dor” e, portanto, poderia integrar os grupos Infelicidade e Pena. Já “pesar” como verbo pode significar “calcular o peso aproximado de” ou “ponderar”, não tendo carga emocional, ou pode significar “provocar desgosto, arrependimento” e “causar tormento”, integrando também o grupo Infelicidade. Os Exemplos 30 a 33 ilustram sentidos emocionais e não emocionais do verbo “pesar”.

Exemplo 30: par=AH-História_da_Inquisição_III-6: No meio dos furores da intolerância, o remoto e o impérvio de alguns distritos que, de ordinário, ainda hoje como que esquecem, para o bem e para o mal, na vida administrativa do País, não eram obstáculo para a mão de ferro da tirania ir lá **pesar** duramente sobre a raça que, porventura, esperava nesses distritos montanhosos e agrestes para obter o esquecimento de um rei fanático e de uma corte hipócrita .

Exemplo 31: par=EQ-Prosas_Bárbaras-281: Jerusalém devia **pesar** à alma delicada e aspiradora do Mestre .

Exemplo 32: [F940116-088](#)-809: Ashrawi -- Acho que ele está sob muita pressão; há muitos fatores que ele deve **pesar**, não é uma situação fácil.

Exemplo 33: [F940210-206](#)-1472: Entre os antílopes africanos está o gunga o maior da espécie, que pode **pesar** até 900 quilos.

A palavra “sentido” é encontrada em três classes gramaticais diferentes. Como substantivo pode significar “direção”, “significado”, “bom senso”; como adjetivo, “magoado”; e como verbo flexionado – cujo lema é “sentir” – “experimental” ou “ser sensível a”. Desses significados, o substantivo não teria carga emocional e o verbo, que é bastante polissêmico, pode descrever emoção como em “sentiu muito por tê-la ofendido” ou não ter sentido emocional, como em “sentiu o joelho na partida anterior”. De qualquer forma, o verbo pode também introduzir a emoção que o falante escolhe descrever: no caso de “ele sentiu amor pelo filho”, “sentir” está introduzindo a descrição de uma emoção (por essa característica, ele faz parte do grupo Genérico, que reúne os lemas que fazem referência a emoção de forma inespecífica). Os Exemplos 34 a 39 elencam os usos de “sentido” enquanto substantivo, adjetivo e verbo.

“sentido” como substantivo

Exemplo 34: <p>: A rua Benjamim Constant, que corta todo o bairro no **sentido** Sul-Norte, foi durante muito tempo a única via de acesso para os outros lugares da região (Barreto, Engenhoca, Tenente Jardim) e para o município de São Gonçalo .

Exemplo 35: <p>: A expressão não tem **sentido** jurídico e não pode ser usada como sinônimo de povo.

“sentido” como adjetivo

Exemplo 36: De novo um grito profundamente **sentido** alastrou entre as massas ali presentes -- 25 de Abril sempre !

Exemplo 37: -- Porque a sorte tinha já assentado que vós devíeis sobreviver a ele, talvez para completar uma existência que vegeta entre as luzes e as sombras do mundo, sem experimentar outras impressões que não sejam as que as sombras, não as luzes, despertam -- respondeu o jovem negociante em tom **sentido** .

“sentido” como verbo (uso não emocional e emocional)

Exemplo 38: <p>: De acordo com o meteorologista Christopher Castro, do CPTEC, o calor que vinha sendo **sentido** na cidade se deve ao fato de que várias massas de ar frio foram para o oceano antes de atingir São Paulo. "

Exemplo 39: par=AG-Flores_sem_Fruto-1: Quanto dera eu por ver e sentir como via e sentia quando pensava pouco e **sentia muito** !

A palavra “obrigado” pode aparecer como interjeição, expressando “gratidão” e pertence ao grupo Gratidão do Emocionário, ou como a flexão do verbo “obrigar”, como sinônimo de “forçar”, que indiretamente pode descrever o sentimento de “insatisfação”, como ilustram os Exemplos 40 e 41.

Exemplo 40: par=fut-BR-JSports-50-39053: Não há um jogador do Bangú que não se sinta **obrigado** a esse esforço .

Exemplo 41: par=t686-4: Em 1905 as partículas de luz voltaram quando o físico suíço Albert Einstein se viu **obrigado** a introduzir a noção de «pacote» de luz (fotão) para descrever o arranque de electrões de um metal por luz ultravioleta .

Os exemplos de “querido” variam entre as classes substantivo, adjetivo e a flexão do verbo “querer” como lema. Dentre seus possíveis significados estão “uma pessoa de quem se gosta”, tanto para substantivo como para adjetivo, ou “ter a intenção de”, no caso da forma do particípio. Dos três casos, apenas os dois primeiros são usualmente do campo semântico de emoção, como mostram os Exemplos 42 a 48.

“querido” como adjetivo

Exemplo 42: par=2505: <> Rômulo Antonio Lima pode sentir de perto o quanto é **querido**, outro dia, em que ele comemorava nova idade... </>

Exemplo 43: par=saude-BR-Revista Brasileira-50-112: Como tudo nesse nosso **querido** Brasil, a Casa da Mãezinha oscila entre um extremo a outro .

“querido” como verbo

Exemplo 44: Isto é precisamente o que ele teria **querido**; só os seus admiradores é que sentimentalmente se queixam .

Exemplo 45: par=fut-BR-JSports-70-44808: Didi, na ocasião, só pode ter **querido** preparar o espírito dos juizes para os jogos dos peruanos .

“querido” como substantivo

Exemplo 46: par=FSP941002-093-1669: (Responde gentil, sorrindo) Claro, meu **querido**, pra poder te mandar à merda mais uma vez !

Exemplo 47: <p>: GUEL ARRAES Estréia no Cultura Artística ‘O Burguês Ridículo’, adaptação de peças de Molière por Guel Arraes e João Falcão Nanini traz a SP seu ‘Burguês’ em êxtase DANIELA ROCHA DA REPORTAGEM LOCAL ELE é descrito como um completo idiota que, por isso, acaba se tornando um **querido** de todos.

Exemplo 48: -- Vae, vae, meu **querido**, porque tambem não podes estar sempre aqui enclausurado .

Nos exemplos de “querido” como substantivo, assim como nas linhas de concordância do AC/DC, é possível perceber que a palavra é usada para descrever emoção, mas também já está cristalizada como um pronome de tratamento, sem carga emocional. Novamente fica clara a importância de o contexto ser considerado em casos de variação de sentido emocional. E por isso também a relevância de haver o projeto de anotação semântica das emoções no Emocionário, que é capaz de dar conta de muitos desses casos de variação de sentido emocional de acordo com a classe gramatical a partir da observação do contexto.

4.4 Argumentos do verbo

Diversos lemas carregam diferentes sentidos, sendo alguns fora do campo semântico de emoção, como descrito no item 4.2, “Polissemia”. Em muitos casos, há a dificuldade em detectar diferentes sentidos de forma automática com base apenas na informação morfossintática, pois compartilham da mesma classe gramatical. Dessa forma, uma solução para anotar semanticamente é observar quais são os argumentos do verbo ou procurar por outras pistas sintáticas ou semânticas. Descrevo aqui os lemas “gozar” e “admirar” – este, complementando a discussão da parte b do item 4.2.

O lema “gozar”, que está no grupo “Satisfação”, apresenta diferentes sentidos, inclusive alguns que só são típicos do português brasileiro, como consta nas entradas dos dicionários digitais Caldas Aulete e Priberam, reproduzidas na Figura 3.

The image shows two columns of text from digital dictionaries. The left column is for the lemma 'gozar' and the right column is for 'go-zar - Conjugar'. The 'gozar' entry includes a pronunciation key (go.zar), a verb class (v.), and five numbered definitions in Portuguese, such as 'Possuir ou usufruir (coisas boas, prazerosas ou úteis); DESFRUTAR; FRUIR' and 'Caçoar ou zombar de (algo ou alguém)'. The 'go-zar - Conjugar' entry shows the conjugation pattern (gozo + -ar), the verb class (verbo transitivo), and six numbered conjugation examples, such as 'Tirar gozo ou prazer de (ex.: gostamos de gozar a vida; gozar a calma e o silêncio) = DESFRUTAR, FRUIR'.

Figura 3 - Definições de “gozar” pelos dicionários digitais Caldas Aulete e Priberam, respectivamente.

Dentre os possíveis objetos do verbo, através da observação do corpus recortada nos Exemplos 49 a 52 é possível verificar que quando o objeto do verbo é indireto e se refere a uma pessoa ou parte de uma pessoa tem o sentido de “caçoar ou zombar”. Já quando esse objeto indireto não diz respeito a uma pessoa, significa “estar na posse” ou “desfrutar”.

Exemplo 49: <p>: Das duas, uma: ou os «dogmáticos» e esse diploma estão **gozando** da nossa cara, ou eles no fundo têm alma escolar, no mau sentido do termo .

Exemplo 50: <p>: A irmã do Hugo **gozava** dele .

Exemplo 51: <p>: Cansas-te em abrir caminho para ela e ela é que vai **gozar** da vida, enquanto aí ficas na caixinha de costura .

Exemplo 52: <p>: Como as demais pessoas, os adolescentes **gozam** de todos os direitos fundamentais e sociais.

Nota-se, portanto, que a descrição de emoção com o lema “gozar” varia de acordo com os argumentos do verbo. O “gozar” transitivo direto significa “desfrutar”; já em seu uso como verbo transitivo seguido de objeto preposicional descreve emoções diferentes no português de Portugal ou do Brasil: se o verbo “gozar” é seguido pela preposição “de”, no português brasileiro pode significar tanto “satisfação” quanto “zombar”, dependendo do objeto; enquanto no português de Portugal, para ter o sentido de “zombar”, o verbo deve estar seguido da preposição “com”. Há ainda a possibilidade de se estar falando sobre o orgasmo sexual, no português brasileiro, quando o verbo é intransitivo. Felizmente, neste caso não se faz necessário criarmos regras de anotação para separarmos os usos de “gozar” como o orgasmo sexual ou com o sentido de “deleitar-se”, já que ambos são parte do grupo Satisfação. Caso contrário, somente a pista do tipo de argumento do verbo não seria suficiente para a criação das regras.

Outro exemplo é o verbo “admirar”, estudado por Santos e Mota (2015) e já mencionado na parte “b” da seção 4.2, que descreve “veneração” em sua forma transitiva e “surpresa” quando na forma “admirar-se”, como mostram os Exemplos 53 a 56, retirados do corpus Literateca em janeiro de 2020.

Exemplo 53: par=AF-Sol_de_Inverno-57: Esqueço-me a **admirar** os teus olhos profundos.

Exemplo 54: par=AG-Helena_(AG)-22: Era com efeito um prodígio de beleza e de perfeição aquela flor; e não precisava ser botânico ou florista para a **admirar** com entusiasmo.

Exemplo 55: par=AV-Sermão_da_Terceira_Quarta-feira_da_Quaresma-32: Deixai perguntar ao mundo e **admirar-se** de vos não ver premiado .

Exemplo 56: par=CCB-A_Doida_do_Candal-217: **Admiro-me** que a justiça lhe tenha deixado desenvolver essa funesta bossa !

Esta seção serviu para evidenciar como os tipos de argumentos do verbo podem nos fornecer pistas para a anotação semântica de emoção e que, por vezes, mesmo que a transitividade seja a mesma, o verbo pode descrever ou não emoção dependendo do objeto ao qual faz referência. Nestes casos, a anotação semântica de emoção se torna uma ferramenta essencial para refinar as potencialidades da pesquisa com corpus.

4.5

Palavras de emoção ocorrendo em um contexto de negação

Um dos problemas ao analisar a descrição de emoção na língua (não só no caso do português) é a utilização de frases negativas ou expressões que indicam negação, característica já identificada no português por Maia e Santos (2012), com o par “medo-coragem”, e Santos e Mota (2015), com o par “surpresa-admiração”. No AC/DC, as palavras de emoção estão anotadas com o grupo referente à emoção descrita pela palavra, mesmo que ela apareça negada. Uma comparação entre os estudos de “medo-coragem” e “surpresa-admiração” (Santos e Mota, 2015) reconhece que, mesmo que o par do primeiro grupo tenha opostos lexicais mais óbvios, a negação do lema “admiração”, do segundo par, aparece em quantidades próximas às do primeiro (2015:68-69).

Maia e Santos (2012) mapearam a utilização de negação com verbos do grupo Medo do Emocionário e, ao encontrarem pouco mais de dez por cento dessas construções em relação ao total de entradas, alguns dados chamaram atenção, como a pré-disposição dos verbos “acovardar”, “assustar” e “arrecear” a serem usados em frases negativas, em relação a verbos como “apavorar” e “aterrorizar”. Já em construções negativas com a presença de um substantivo do grupo Medo apareceram em quase vinte por cento do total, sendo cerca de oitenta por cento deles antecedidos pela preposição “sem” (os padrões mais frequentes sendo “sem medo”, “sem receio” e “sem sobressalto”).

A partir desta análise específica sobre o grupo Medo e a utilização das palavras pertencentes a ele, as autoras constataram que falantes da língua portuguesa adotam as mesmas construções gramaticais para referirem-se a medo e coragem, incluindo as expressões metafóricas:

“Os exemplos ‘sem medo’ trazem à tona o status de coragem em oposição ao léxico de medo (...) e, em algumas culturas, fica claro que medo e coragem se relacionam em nível moral. Essas informações tornam ainda mais interessante a percepção de que coragem e medo são similares na língua portuguesa, no sentido de que são expressados usando os mesmos operadores gramaticais e se utilizam das mesmas metáforas.” (Maia e Santos, 2012, tradução minha)

Como a negação em palavras é na grande maioria dos casos marcada por um afixo (que cria um novo lema e é discutido na seção 4.6 desta tese) ou uma outra palavra de negação compondo a frase, a opção segura seria manter a anotação de emoção original das palavras e, caso o pesquisador queira estudar a negação da emoção, as ferramentas de busca e o corpus anotado favorecem

a pesquisa, mesmo que a negação não esteja anotada diretamente nas palavras de emoção.

4.6

Morfologia: prefixos e sufixos em palavras de emoção

Outra característica da língua portuguesa presente na menção de emoção é a utilização de prefixação e sufixação. Apesar de não representar um obstáculo para a anotação, por criar lemas diferentes e que são anotados de forma independente, é necessário um olhar atento do pesquisador, já que nem sempre o prefixo negativo significa uma oposição à palavra original.

Assim como os ditos “pares” de grupos de emoção não necessariamente tem sentidos complementares, também as palavras “confiança” e “desconfiança” não se complementam em sentido, já que “desconfiança” não significa somente “falta de confiança”. O mesmo fenômeno acontece com “esperar” e “desesperar”, quando o oposto de “esperar” seria “não esperar”, e em lemas como “desgosto” e “desgostar”, que não são simplesmente a negação de “gosto” e “gostar” (como também refere Sandmann, 1997), pois passam por acréscimos semânticos e mudança de significado além da prefixação. É necessário, pois, refletir em que grupos devem estar classificados, sobretudo nos casos em que não há um grupo explícito a ser atribuído (palavras como “desrespeito”, “desgosto”, “desencorajar”, por exemplo).

4.7

Mudanças de tempo e modo verbal

O verbo “acreditar” é um caso que deixa transparecer claramente que a mudança no tempo verbal pode implicar na mudança de sentido na descrição de emoção. “Acreditar” faz parte do grupo esperança, e “acreditava” ou “acreditei”, por exemplo, apesar de descreverem esperança, quando usados no passado, descrevem uma esperança por algo que não aconteceu conforme esperado, o que descreveria uma frustração, por exemplo.

Já em sua utilização entre vírgulas (constituindo uma oração intercalada), “acreditar” pode carregar os sentidos de (i) confiar, quando usado no imperativo;

(ii) ter algo como provável, quando nas primeiras pessoas do singular e do plural do indicativo; e (iii) ter algo como verdadeiro, quando nas terceiras pessoas do singular e do plural do presente ou pretérito imperfeito. Mesmo com três sentidos diferentes, ainda sim “acreditar” faz menção aos sentidos do grupo Esperança, como confirmam os Exemplos 57 a 67.

(i) Confiar:

Exemplo 57: <p>: Nas duas noites seguintes, **acredite**, nenhuma única pena se repete .

Exemplo 58: par=FSP940509-093-1157: A porta de entrada dos «teens» no universo das armas é, **acredite**, a aquisição de revólveres e metralhadoras de plástico .

Exemplo 59: par=14573: Mas o treinador tetracampeão mundial, **acreditem**, está fazendo milagres.

Exemplo 60: par=FSP940123-009-41: Além de aumentar o número de cassações, deve-se estimular as demais CPIs sobre Empreiteiras, financiamentos eleitorais e CUT a CPI da CUT terá o dom de trazer à tona possíveis desvios da burocracia e da praga do corporativismo, o que , **acreditem**, vai acabar ajudando não apenas o PT mas a democracia

(ii) Ter algo como provável:

Exemplo 61: par=5634: E a Feira do Livro tem os seus créditos firmados dentro e fora da Cidade e , **acredito**, fora das fronteiras pelo menos na vizinha e irmã Galiza

Exemplo 62: par=saude-PT-Saude Lar-50-35969: Basta , **acredito**, a enumeração acima, como prova sugestiva de que é sempre necessário rodear a piscina de cuidados higiénicos .

Exemplo 63: par=fut-PT-Record-70-25053: Ao assistente , **acreditamos**, parece ser muito fácil meter a bola na baliza nas condições em que por vezes, antontem, Lourenço se apresentou .

Exemplo 64: par=7190: Tendo em vista a quantidade e qualidade dos produtos, dassim como a eficiente e criteriosa cobertura publicitária , **acreditamos**, em princípio, no sucesso de vendas no varejo, uma vez que se trata, também, de produtos de pouca divulgação em nossa região .

(iii) Ter algo como verdadeiro

Exemplo 65: par=ext25569-pol-92a-1: Quem as possuísse , **acreditavam**, teria garantida ao mínimo a provação do purgatório .

Exemplo 66: Para o actual clima de confiança Armando Morais invoca ainda o muito trabalho já realizado nesta fase de pré-campanha, traduzido em inúmeras iniciativas, trabalho este que, **acredita**, irá dar os seus frutos .

Exemplo 67: <p>: E a resposta a iniciativas nesse sentido , **acredita**, está sendo extremamente positiva .

É importante ressaltar que os verbos que se incluem no caso de mudança de sentido de acordo com o tempo verbal podem ser inseridos em diferentes grupos do Emocionário de acordo com o sentido emocional descrito em relação à flexão temporal. O corpus anotado garante a qualquer pesquisador a elasticidade de buscar por flexões temporais e número-pessoais, caso tenha

interesse em estudar essa relação entre emoção e mudanças de tempo e modo verbal mais a fundo e considerar possíveis variações, não só de verbos, mas em todas as classes gramaticais.

4.8 O peso emocional das palavras de emoção

A carga semântica emocional de algumas palavras é mais forte que de outras. O verbo “pedir”, por exemplo, não descreve emoção por si só, ao passo que os verbos “suplicar” ou “implorar”, significam “pedir com súplica, humildade ou mesmo em uma situação de desespero”. Ou seja, “suplicar” e “implorar” funcionam como a ação de “pedir” carregada com uma intenção emocional, como nos Exemplos 68 a 73.

Exemplo 68: id="A_Mortalha_de_Alzira Prosa:romance AA 1894 naturalismo_realismo_romantismo masc ": Sua alma sangrava ainda, pedindo mais sacrifícios, e ele caía de joelhos, arranhando as carnes do peito com as unhas, e **suplicando** a Deus que lhe inspirasse um meio de resgatar-se, completamente, aos olhos da sua própria consciência envergonhada .

Exemplo 69: id="Mano Prosa:romance CN 1924 masc ": De repente, em ímpeto de desespero, ajoelhando-se, com os braços estendidos sobre a mesa, de mãos postas, **suplicava**...

Exemplo 70: id="O_monstro_e_outros_contos Prosa:conto HC 1932 masc ": Se, depois, recusardes o que vos **suplico**, morrerá, para mim, a última esperança de paz e libertação !

Exemplo 71: par=4: Mexeu de um lado para o outro com a cabeça, **implorando** com seus olhos para que tirassem aquele tubo, e a deixassem morrer em paz .

Exemplo 72: par=saude-BR-Revista Brasileira-50-173: Homens e mulheres **imploram** por um leito, às vezes se deixando ficar à porta dos Hospitais de Isolamento, indiferentes ao tempo, aguardando por uma vaga como quem espera pela sorte grande .

Exemplo 73: par=FSP940605-068-986: Os registros de Walla Walla incluem descrições de presos sendo parcialmente decapitados, por terem sido pendurados de alturas grandes demais, ou casos em que os condenados **imploraram** ao carrasco que os erguesse e deixasse cair novamente .

O verbo “pedir” pode descrever mais ou menos emoção em contextos nos quais é combinado com outras palavras de emoção, mas não de forma isolada. Por isso é importante analisarmos a língua em contexto. A Figura 4 mostra o que os falantes de português pedem no corpus OBras:

Procura: [lema="pedir"] @[pos="N"]
Distribuição de lema
Corpo: OBras v. 8.4

542 casos.

Distribuição

Houve 200 valores diferentes de lema.

licença	93	explicação	8	alvissara	5	indenização	2	elogio	2
desculpa	28	informação	7	justiça	4	ficha	2	nova	2
esmola	22	favor	7	compaixão	4	repetição	2	cerveja	2
notícia	20	coisa	7	amor	4	pouso	2	parati	2
água	15	silêncio	7	segredo	4	vênia	2	charuto	2
explicação	14	café	6	meça	3	satisfação	2	apoio	2
demissão	14	conta	6	permissão	3	juramento	2	esclarecimento	2
conselho	11	pão	6	proteção	3	alívio	2	providência	2
dinheiro	11	fogo	6	descanso	3	privilégio	2	agasalho	2
misericórdia	10	pousada	5	tempo	3	emprego	2	repouso	2

Figura 4 - Distribuição de complementos do verbo “pedir” no corpus Obras.

A Figura 4 mostra os cinquenta primeiros lemas de uma distribuição total de duzentos. Por esse recorte, já é possível perceber a carga emocional diferente entre os complementos de “pedir”: “pedir notícias” pode significar “preocupação”, mas o contexto das frases deve ser analisado para chegarmos a essa conclusão. Já “pedir misericórdia” tem sentido muito mais emocional, pois “misericórdia” é uma palavra que dá menos abertura à variação de sentidos. Existe, portanto, um gradiente entre usos que descrevem claramente emoções e outros que não são emocionais.

Como dito inicialmente, o principal objetivo deste capítulo era encontrar caminhos para desafios que permeiam a análise de menção de emoção, como qual caminho utilizar para confirmar se uma palavra descreve emoção; em que casos a palavra pode ser considerada uma descrição de emoção e como agrupar as palavras que descrevem emoção. Após elencar as características e pontos que merecem atenção, foi possível perceber que uma das saídas para contornar esses desafios é a observação do contexto para que então sejam criadas regras de anotação. Logo, reconheço e discuto neste capítulo a característica das possíveis mudanças e variações entre sentido emocional ou não emocional no recorte de palavras/lemas oferecidas pela riqueza da língua. Este capítulo descreve as características e desafios da análise de menção de emoção também como forma de introduzir a ampliação e validação do Emocionário através de diferentes explorações no AC/DC detalhadas no capítulo 5 desta tese.

5 Ampliação e validação do Emocionário

Este capítulo trata das explorações em corpora feitas com o objetivo de ampliar e validar o conteúdo do Emocionário. Inicialmente, discorro sobre a busca por padrões léxico-sintáticos de Hearst (1992) e sobre como ela pode servir de inspiração para o desenvolvimento e a busca por padrões léxico-sintáticos no português, visando a enriquecer a análise de corpus. Um tipo de busca no corpus a partir de padrões léxico-sintáticos foi sugerido por Hearst no início dos anos noventa. O método consiste em extrair relações de hiponímia de grandes quantidades de texto de forma automática. Para a busca, a autora propôs padrões léxico-sintáticos que fossem frequentes, aparecessem em diferentes gêneros textuais, revelassem a relação sintática desejada (nesse caso, a hiponímia) e fossem facilmente reconhecidos (1992:540). Dessa forma, seria possível abranger expressões variadas nas quais a relação de hiponímia pudesse aparecer.

Dentre os objetivos de aplicar essa abordagem estava a desassociação de projetos de extração de informação lexical dos dicionários em formato digital, que são limitados na medida em que neles sempre há um número fixo de entradas. Outros objetivos eram aprimorar a expansão de sinônimos na extração de informação, encontrar cadeias de orações que estão semanticamente relacionadas e auxiliar programas de PLN, o que tornaria a aquisição automática de léxico mais atraente para o reconhecimento de relações baseadas em padrões. Um corpus é uma fonte de informações sobre a língua na qual ele foi escrito: ao buscar por relações lexicais específicas comumente usadas, somos capazes de explorar padrões e observar e descrever aspectos que passam despercebidos em uma leitura linear.

Essa busca também se mostra útil para a confirmação de padrões léxico-sintáticos na língua portuguesa. Como mostra a Figura 5, a expressão relativamente simples [pos="N.*" & func ="SUBJ>"] "é" "um" "sentimento", os dados do corpus Literateca no AC/DC se fazem relevantes para esta pesquisa.

Procura: [pos="N.*" & func="SUBJ>"] "é" "um" "sentimento"
Distribuição de lema
Corpo: Literateca v. 7.9

15 casos.

Distribuição

Houve 10 valores diferentes de lema.

amor	5
ódio	2
admiração	1
amôr	1
amizade	1
bondade	1
compaixão	1
inveja	1
pátria	1
tristeza	1

Figura 5 - Resultados da busca pelo padrão léxico-gramatical “[pos="N.*" & func="SUBJ>"] "é" "um" "sentimento”.

Além de a maior parte dos lemas encontrados já serem palavras do léxico de palavras de sentimento criado anteriormente, o dado mais pertinente da lista mostra justamente que “pátria” é uma das palavras à qual os falantes de português recorrem para descrever sentimentos, apesar de não fazer parte do campo lexical de emoções: “*par=JN-Camões-8: A pátria é um sentimento enérgico, desinteressado, benéfico, mesmo quando é um fanatismo.*”

O pensamento de Hearst (1992) em relação à língua e a relações semânticas entre orações pode ser trazido para o português, pois termos que coexistem em um mesmo campo lexical estão relacionados semanticamente, ainda que não apareçam diretamente juntos no corpus. Além disso, relações que conectem semanticamente *injury* e *broken bone* provavelmente não seriam encontradas em léxicos ou dicionários, bem como seus equivalentes em português (“lesão” e “osso quebrado”, respectivamente).

No entanto, algumas informações se contrapõem entre o inglês e o português. Para a pesquisa de Hearst (1992:541), foram levados em conta apenas alguns modificadores, como adjetivos quantificadores, como *some* e *many* (alguns e vários, respectivamente). Nesse caso, a equivalência para o português: “*many broken bones*” e “muitos ossos quebrados” seriam analisados em níveis iguais. Já a pesquisa em inglês poderia abarcar também advérbios, como *completely* (“completamente”) e *really* (“realmente”), diferentemente de

uma pesquisa semelhante em português. No inglês, o modificador continuaria no começo da locução, “completely broken bones”, enquanto em português, o modificador entraria no meio do sintagma nominal, “ossos completamente quebrados”, criando um obstáculo na obtenção de um padrão léxico-gramatical como esse. Ainda sim, o método criado por Hearst é relevante para inspirar a obtenção de padrões que façam sentido na língua que está sendo pesquisada.

Ressalto aqui que esta pesquisa argumenta a favor da necessidade de adaptação e desenvolvimento de métodos próprios para análise linguística específicos para o português. Métodos e abordagens já existentes para análise linguística em outras línguas podem e devem ser usados como inspiração e para que, a partir da experiência compartilhada por pesquisadores de outras línguas, pesquisadores da língua portuguesa tenham em mãos recursos que levem em conta as particularidades da nossa língua, obtendo resultados ainda mais significativos. A seção 5.1 descreve como ocorreu a pesquisa pelo padrão “sentimento de N”; explica como foi feito o tratamento dos lemas com menos de três ocorrências no corpus (cauda longa) e os resultados da busca, incluindo a descoberta de campos semânticos que tangenciam o das emoções e são usados, no contexto da pesquisa, para descrever emoções. A seção 5.2 explica o processo da pesquisa por menção de emoção através de expressões com o corpo humano, resultante da exploração da anotação de palavras do corpo humano do projeto Esqueleto.

5.1 “Sentimento de N”

A pesquisa pelo padrão léxico-gramatical “sentimento de N” (Ramos e Freitas, 2019) aconteceu a partir de uma ampla varredura nos corpora do AC/DC que, à época, já contava com mais de um bilhão de palavras. Para elaborarmos o léxico, partimos da expressão de busca “sentimento de N”²¹ no corpus OBRas (Santos et al., 2018), por meio da interface AC/DC, criada e mantida pela Liguatca. Trata-se de uma abordagem simples, inspirada nos padrões de Hearst (1992), mas que nos assegura de que a palavra em questão está sendo entendida, pelo falante, como uma palavra de sentimento, e dessa maneira nos desviamos da polêmica sobre o que é um sentimento na língua.

²¹ Especificamente, a expressão de busca utilizada foi [lema="sentimento"] [lema="de"] [pos!="V"]* @[pos="N" & func="P<"] within s

A escolha de iniciarmos a análise pelo corpus OBras se deveu a dois principais motivos: trata-se de um corpus de obras literárias, e partimos do princípio de que a literatura seria um espaço propício para procurar emoções, e por seu tamanho²². No entanto, diferentemente do que esperávamos, a busca não forneceu diretamente tipos de sentimento. Após a análise de diversos casos, distribuímos os lemas encontrados em quatro grupos:

- **Grupo 1:** O “N” corresponde a sentimentos que consideramos convencionais, como *culpa*, *medo* etc.
- **Grupo 2:** O “N” corresponde a maneiras não convencionais de falar sobre sentimentos.
- **Grupo 3:** O “N” não se refere a um sentimento, mas ao possuidor/experenciador do sentimento.
- **Grupo 4:** O “N” introduz um modificador do sentimento. Isto é, o “N” não nomeia um sentimento, apenas atribui a ele qualidades (mostrando que o padrão é polissêmico).

O Quadro 4 apresenta exemplos do corpus para cada um dos grupos mencionados.

Grupo 1	<i>id="Iaiá_Garcia Prosa:romance MdA 1878 romantismo"</i> : Luís Garcia não pôde furtar-se a um sentimento de pena , ao vê-lo entrar fardado e prestes a seguir para o Sul
Grupo 2	<i>id="A_semana Prosa:crônica MdA 1892 "</i> : Atentai, mais que tudo, para esse sentimento de unidade nacional , que a política pode alterar ou afrouxar, mas que a arte afirma e confirma, sem restrição de espécie alguma, sem desacordos, sem contrastes de opinião
Grupo 3	<i>id="O_Matuto Prosa:romance FT 1878 realismo_regionalismo_romantismo"</i> : 'Tais eram as idéias e os sentimentos de d. Damiana
Grupo 4	<i>id="A_Alma_Encantadora_das_Ruas Prosa:crônica JdR 1807 "</i> : Esse sentimento de natureza toda íntima não vos seria revelado por mim se não julgasse, e razões não tivesse para julgar, que este amor assim absoluto e assim exagerado é partilhado por todos vós

Quadro 4 - Exemplos de tipos de sentimento conforme os grupos de análise.

Foram encontrados 120 lemas diferentes a partir dessa abordagem. Para um léxico de sentimentos, nosso interesse está nos grupos 1 e 2. Nosso foco está em descartar da lista aquilo que claramente não se refere a um tipo de sentimento. Assim, uma outra maneira de olhar para os grupos é vê-los simplesmente como *sentimento* (1 e 2) vs *não-sentimento* (3 e 4).

Em seguida, após termos clareza sobre a maneira de lidar com os resultados inesperados, repetimos o procedimento, dessa vez tomando como objeto todos os corpora disponíveis no projeto AC/DC – note-se que, com isso,

²² 5,7 milhões palavras em setembro de 2019

estamos lidando com aquilo que é materializado como sentimento na língua portuguesa nas variantes brasileira, portuguesa e moçambicana, e em diferentes gêneros e recortes temporais.

A busca em todos os corpora (excluindo-se o OBRas) do AC/DC resultou em 2060 lemas diferentes. O ponto de corte da análise dos lemas foi a frequência acima de 3, o que diminuiu o número de lemas para 853. Alguns lemas foram excluídos por serem erros de digitação ou pré-processamento (por exemplo, “germanidade” também aparecia como “germanidade6”), então excluímos as duplicidades, e 86 lemas já haviam sido analisados na etapa do corpus OBRas. Além dessas exclusões, também agrupamos oito lemas por representarem apenas variações ortográficas (por exemplo “afeto” e “afecto”) e, com isso, ficamos com 724 lemas para a análise.

Os resultados estão na Tabela 3. Indicamos com “sema = emo” todos os lemas que já apareciam anotados no AC/DC como pertencentes ao campo semântico da emoção. Do mesmo modo, “sema ≠ emo” indica as palavras que nós consideramos palavras de sentimento convencionais (Grupo 1) ou pouco convencionais (Grupo 2), mas que não haviam sido anotadas como tal no AC/DC.

		Corpus		
		OBRas	Todos exceto OBRas	Todos
Grupo 1	sema = emo	33 (62%)	76 (35%)	109 (40%)
	sema ≠ emo	20 (38%)	141 (65%)	161 (60%)
Grupo 2	sema = emo	0%	26 (6,1%)	26 (5,5%)
	sema ≠ emo	49 (100%)	398 (93,9%)	446 (94,5%)
Grupo 3		11	59	70
Grupo 4		07	24	31

Tabela 3 - Resultados da análise levando com conta a anotação do AC/DC.

Graças à camada anterior de anotação semântica de emoções, também foi possível compararmos a eficácia de nossa abordagem (e indiretamente, podemos avaliar a anotação semiautomática já presente no AC/DC). Os dados mostram que o padrão “sentimento de N” é confiável, levando a 90% de palavras de sentimento. Quando comparamos com a anotação do AC/DC, vemos que, no

corpus OBRas, apenas 62% das palavras que consideramos palavras de “sentimentos convencionais” já estavam identificadas como palavras do campo semântico das emoções/sentimentos. Quando vamos para todo o material, apenas 40% de tudo o que consideramos “sentimentos convencionais” já continha anotação de emoção. Alguns exemplos de palavras do grupo de sentimentos pouco convencionais que encontramos e que estavam com “sema=emo” são “autoestima”, “decoro” e “heroísmo”.

Uma explicação para essa divergência está na própria anotação do AC/DC. Expressões do tipo “sentimento de N” contêm a indicação de uma etiqueta de emoção genérica (“emo:gen”) na própria palavra “sentimento”, e o substantivo especificador do sentimento, por outro lado, não recebe sema algum. Além disso, mas de maneira não surpreendente, no OBRas, nenhuma das palavras que consideramos palavras de sentimentos pouco convencionais já estava anotada como palavra do campo semântico das emoções/sentimentos. Por outro lado, quando consideramos o AC/DC completo, temos que 5% das palavras de sentimento que consideramos pouco convencionais já estavam anotadas como pertencente ao campo semântico das emoções, o que sugere que fomos rigorosas com a ideia de convencionalidade. Quanto à ampliação do léxico do Emocionário, decidimos atribuir a um grupo de emoção todas as palavras dos Grupos 1 e 2 – essa distribuição está detalhada no Anexo 1 desta tese. A partir dela, como uma das contribuições da exploração do padrão “sentimento de N”, foram criados os grupos Outra e Ausência para o Emocionário, que marcam respectivamente casos que se referem à emoção mas não conseguimos encaixar em nenhum dos grupos já existentes e casos que se referem à inexistência de emoção.

Ao longo da análise dos lemas, nos chamou a atenção a distribuição das palavras de sentimento por campos lexicais pouco mencionados na literatura sobre emoção: (1) palavras de sentimento relacionadas à ideia de *pátria*; (2) palavras de sentimento relacionadas à *religião* e *espiritualidade*; (3) palavras do campo semântico de *família* e *parentesco* usadas para expressar sentimento. Os referidos lemas estão no Quadro 5. Destes, apenas “ufania” já estava anotado semanticamente no AC/DC, com a etiqueta “sema=emo:orgulho”.

A distribuição dos lemas por campos lexicais promove reflexões interessantes. Grande parte das palavras de sentimento usadas nos corpora estão contidas outros campos lexicais que não o das emoções – nacionalismo, espiritualidade e parentesco. Além disso, lemas como “brasilidade”, “patriotismo” e “pertencimento”, que aparecem no Quadro 5, podem carregar mais de um

sentimento, como por exemplo “orgulho”, “amor”, “(in)satisfação” e “(in)felicidade”.

<p>Campo lexical 1: lemas referentes ao nacionalismo. Sentimento de...</p> <p>americanismo – anticomunismo – antilhanía – anti-lusitanismo – bairrismo – brasilidade – cidadania – civismo – democracia – germanidade – germanismo – identidade – italianidade – mineiridade – nação – nativismo – nacionalização – nacionalidade – nacionalismo – origem – pátria – patriota – patriotismo – pertença – pertencimento – povo – raça – ufanía – ufanismo – união – xenofobia.</p> <p>Exemplos: <i>“É que o sentimento de brasilidade, de patriotismo, nato do brasileiro, impera”</i> <i>“Eu fui ao diante dele, afirmando que a adoção de uma nacionalidade é ato político, e muita vez pode ser dever humano, que não faz perder o sentimento de origem nem a memória do berço”</i> <i>“Essa instituição por ele pregada e que se fez realidade entre nós precisa ser conservada sempre, não só por ser ela a defesa do território, como ainda por servir de meio eficaz e rápido para civilizar o caipira, dando-lhe o sentimento de pátria e ensinando-lhe amar e compreender as cores e a história da nossa Bandeira”</i></p>
<p>Campo lexical 2: lemas referentes à espiritualidade. Sentimento de...</p> <p>abstenção – adoração – católico – cobiça – comunhão – devoção – doação – fanatismo – fé – idolatria – perdão – religião – religiosidade</p> <p>Exemplos: <i>“Helena deixou-se cativar desse sentimento de abstenção e elevação; se alguma dor ou remorso a pungia, esqueceu-os, por um minuto ao menos, entre aquelas paredes desataviadas, diante de um padre, entre uma imagem de Jesus e as obras vivas do Criador”</i> <i>“Se escrevo sobre o médium, não é por nenhum sentimento de idolatria, mas por reconhecimento ao trabalho de um companheiro que abdicou de si mesmo para servir a causa que abraçou”</i></p>
<p>Campo lexical 3: lemas referentes a família e parentesco. Sentimento de...</p> <p>avó – irmandade – irmão – mãe – maternidade – orfandade – pai – paternidade – viúvo</p> <p>Exemplos: <i>“Ao contrário do que se pensa, o sentimento de maternidade não é natural, mas sim uma construção social e cultural, como já foi mostrado nos capítulos anteriores”</i> <i>“Carlos, o marido de Edemeia é a personagem que completa junto com Carvalho, pai da triângulo da contenda e reproduz a figura do ser angustiado e dividido entre a manutenção vigor e poder reservadas ao sexo masculino, e o sentimento de pai e de esposo”</i></p>

Quadro 5 - Campos lexicais identificados no léxico de emoções.

5.1.1 Tratamento da cauda longa do “sentimento de N”

A Lei de Zipf, popularizada na década de 1940, descreve que as relações entre frequência e posição em um ranking decrescente de dados tende a ser inversamente proporcional – ou seja, quanto mais alto na posição do ranking, maior sua frequência – sendo a frequência da primeira aproximadamente duas vezes maior que a da segunda posição, três vezes maior que a da terceira, e

assim por diante, até que para as posições médias e finais do ranking, há uma grande quantidade de dados com baixa frequência em corpus, que é considerada a “cauda longa” da análise. A cauda longa do “Sentimento de N” era composta por 1207 palavras que apareciam apenas uma ou duas vezes no corpus. A estratégia adotada para analisá-la foi o tratamento manual da lista, que aconteceu em três etapas. A primeira foi descartar o “lixo” – isto é, palavras que continham erros de digitação ou que estavam mal escritas. Destaco que esses lemas já haviam sido contados com a escrita correta na lista inicial, como “egoísmo”, que aparece na cauda longa sem o acento, e “pertencimento”, que aparece digitado como “pertecimento” e “pertenciamento”. Os No Quadro 6 aparecem os exemplos dos lemas mencionados em contexto.

<p>Lema: egoísmo “A ausência da compreensão dessa dinâmica pode trazer para as atuais gerações um sentimento de egoísmo e descompromisso com o futuro.” “e punham no triste abandono daquele quarto uma melancolia dura, um áspero sentimento de egoísmo; alguma coisa da indiferença dos que vivem pelos que se vão meter silenciosamente dentro da terra.”</p>
<p>Lema: egoismo “Quando se é leal e sincero para com todos; quando se não abrigam no peito sentimentos de egoismo e de torpe má fé, a, que vela pelos bons, enviamos sempre alguém que nos compreende e nos vale nos lances mais desesperados da nossa vida...”</p>
<p>Lema: pertencimento “Para constituição do vínculo com o grupo, é imprescindível o sentimento de pertencimento, principalmente naquelas que possuem características particulares.”</p>
<p>Lema: pertenciamento “A seleção pode realizar a tarefa, exercida por Senna há até pouco tempo, de propiciar na população um sentimento de pertenciamento, de se sentir coletivo, de estar junto com mais gente, de executar um projeto comum, de ser nação”</p>
<p>Lema: pertecimento “A Educação Ambiental pode resgatar as sensações valorativas para que as subjetividades individual e coletiva criem um sentimento de pertecimento à natureza, de um contado íntimo com a natureza para perceber a vida em movimento de equilíbrio / desequilíbrio, organização / desorganização, vida / morte, o belo e o bom nela contidos.”</p>

Quadro 6 - Exemplos de lemas com e sem erros de digitação.

Nesta etapa, também chamaram a atenção doze lemas por serem palavras importadas de outras línguas, como “pueblo” e “ghetto”, ilustrados no Quadro 7.

<p>Lema: pueblo “«Quis transmitir o sentimento de ` pueblo`, o amor ` campesino` pela terra que senti quando era miúdo.”</p>
--

Lema: ghetto

“E Jesse Jackson é evidentemente o último dos líderes que o anti-tribalismo precisaria: é um puro caçador de votos, que estimula o **sentimento de «ghetto»** dos negros americanos, em vez de promover o seu orgulho de cidadãos americanos.”

Quadro 7 - Exemplos de lemas importados de outras línguas.

É curioso destacar que, ao serem traduzidas para o português (“cidade” e “gueto”, respectivamente) e inseridas no contexto das respectivas frases, possuem lemas correspondentes em português que também apareciam dentre os resultados da busca pelo padrão “sentimento de N”. Ao final desta primeira etapa, foram retirados aproximadamente 10% dos lemas da cauda longa, que ainda sim permanecia com pouco mais de mil palavras.

A segunda etapa consistiu em agrupar essas palavras em palavras de sentimento ou palavras que não são de sentimento. Este passo ocorreu de forma diferente da pesquisa pelo padrão “sentimento de N”: independentemente de serem convencionais ou não convencionais, após a análise caso a caso, todas as palavras consideradas de sentimento na análise do contexto foram agrupadas em um mesmo arquivo e encaminhadas para o passo seguinte. Os lemas que não eram de sentimento, por sua vez, constituíam quase 50% da cauda longa e foram descartados nesta etapa. Alguns desses lemas foram descartados como não sendo palavras de sentimento por três principais razões, exemplificadas no Quadro 8. A primeira delas é que alguns dos lemas apenas apareceram como resultado por falta de maior especificação na expressão de busca, como mostram os lemas “nível”, “proporção” e “questão”. A segunda razão ilustrada através de entradas com os lemas “enriquecimento” e “fase” é a polissemia da palavra “sentimento” introduzindo um sintagma nominal e, logo, não fazem parte do padrão “sentimento de N”. A terceira razão para os lemas serem desconsiderados da análise é a polissemia do padrão, ou seja, trata-se de lemas que apontam o possuidor/experienciador do sentimento, como os exemplos com os lemas “filósofo”, “fotógrafo” e “gestante”, e não de palavras que descrevem sentimentos.

<p>Lemas: nível, proporção e questão</p> <p>“Em «Amsterdam Global Village», há o sentimento de um movimento circular no nível da água e também um movimento aéreo, quando voamos para diferentes países.”</p> <p>“De um lado, acentua-se o sentimento de medo e insegurança coletivos, na mesma proporção em que as políticas de segurança e justiça convencionais revelam-se ineficientes para garantir a lei e a ordem.”</p> <p>“As manifestações, prevê Rahimullah Yusufzai, vão continuar porque «o sentimento anti-americano cresceu nos últimos meses», às custas de um sentimento de injustiça no caso dos testes nucleares ou de uma série de questões menos importantes mas muito abrasivas.”</p>
<p>Lemas: enriquecimento e fase</p> <p>“A perestroika vem incutir um sentimento de enriquecimento rápido -- facilitado pela confusão política, económica e mesmo legislativa gerada pela introdução das reformas -- e sobretudo a qualquer preço.”</p> <p>“Derivam estas formas comuns da expressão do sentimento de fases sociais análogas nas raças superiores; o lirismo nasceu da idealização da vida doméstica, e desenvolveu-se na estabilidade do trabalho pastoral ou agrícola, como se vê nos epitalâmios ou cantos nupciais, nas cantigas do berço ou nanarisma, nos cantos funerais ou nénias, nos cantos das segadas e das vindimas.”</p>
<p>Lemas: Filósofo, fotógrafo e gestante</p> <p>“Ele dizia que a única coisa que sabia era que não sabia de nada, esse o verdadeiro sentimento de um filósofo”</p> <p>“A visita à exposição do Palácio das Galveias permite compreender o sentimento de frustração de muitos fotógrafos da imprensa: a qualidade estética, e técnica de certas imagens fazem-nos pensar no modo como «vemos» aquelas mesmas fotografias no contexto da página do jornal...”</p> <p>“A ultra-sonografia obstétrica e suas Implicações na relação mãe-feto: Impressões e sentimentos de gestantes com e sem diagnóstico de anormalidade fetal.”</p>

Quadro 8 - Alguns lemas que foram descartados ao longo da análise da cauda longa.

Finalmente, a terceira etapa do tratamento da cauda longa consistiu na distribuição das palavras remanescentes em grupos do Emocionário. Apesar de muitas das palavras de sentimento extraídas do padrão “sentimento de N” não serem usadas para descrever sentimento, a descrição do próprio léxico do Emocionário explica que suas listas são compostas por palavras que possuem sentido de emoção em algum contexto. Além disso, não pareceu interessante que esses resultados fossem descartados de uma possível análise sobre categorização de emoções. Logo, um caminho seria: uma vez que a expressão seja atribuída a um grupo do Emocionário, adicionamos a ela a etiqueta semântica com uma regra específica para a expressão “Sentimento de N”. Essa adição possibilita que essas palavras sejam levadas em consideração em análises futuras.

Sendo assim, alguns dos lemas encontrados com o padrão “sentimento de N” foram adicionados ao Emocionário dentro do contexto da expressão (“sentimento de inatividade” ao invés de somente “inatividade”) e incluídos na anotação semântica de emoção no padrão “sentimento de N”. O Quadro 9 mostra uma lista com dez desses lemas.

Grupo AUSÊNCIA:Lema: inatividade

As crianças agredidas precisam repetir a situação da agressão, em uma tentativa de obter o controle e dividir o «poder» com o agressor, a partir de uma identificação com o mesmo, para neutralizar os **sentimentos de inatividade e** aniquilamento associados aos maus-tratos.

Lema: anestesia

O pavor ` anestesia ', afirmam vítimas PRISCILA LAMBERT da Reportagem Local Apesar do crescente pavor, as vítimas de atos de violência sexual são tomadas por um **sentimento de» anestesia \ "**, um tipo de defesa natural para diminuir o sofrimento durante o estupro. "

Grupo OUTRAS:Lema: amputação

Stanley Dance, um dos biógrafos da banda, traduziu bem o **sentimento de amputação** que representou a saída do saxofonista:

Lema: viscosidade

Mas se o **sentimento de viscosidade** deve então ser pensado com relação ao projeto do ser no mundo, vale lembrar uma vez mais que esse projeto, pelo qual o homem humaniza o mundo através do trabalho, equivale, como se viu, ao plano de sua apropriação instrumental.

Grupo INSATISFAÇÃO:Lema: agitação

Moskovitz e Byrd17 definem algumas características como: ato significando salvação do paciente ou do mundo, ausência de arrependimento, citação do versículo bíblico Mateus 5:29, geralmente em pacientes com diagnóstico de episódios psicóticos agudos acrescentam considerações sobre a presença de sentimentos de culpa geralmente associados a questões sexuais nesses pacientes, além de delírios relacionados a uma religiosidade fanática e a **sentimentos de agitação** e de ansiedade intensas seguidos por alívio e tranquilidade após a realização do ato autolesivo.

Lema: rebelião

Meu pronunciamento, Sr. Presidente, não traduz nenhum sentimento pessimista, nenhum **sentimento de rebelião**, conquanto tenha teor profético, sem presunção.

Grupo VERGONHA:Lema: nudez

Este **sentimento de completa nudez** e de subjetividade exposta e sem defesa, está no \ " coração da vergonha \ "

Lema: vexame

O **sentimento de desilusão, abandono e até de vexame** manifestado por este ex-sargento ajudante pela forma como a Marinha está a tratar os seus antigos servidores, é aliás partilhado por quase todos os residentes naquela zona habitacional, integrada no perímetro militar da base.

Grupo ÓDIO:Lema: revanche

A figura do imigrante português será um espécie de significante tenso ou ideologicamente carregado, através do qual os escritores brasileiros (sobretudo no século XIX) irão projectar um ressentimento colonial persistente e complexo, nassim como um **sentimento de revanche** relativamente à antiga «mãe pátria», que é vista por estes escritores como uma nação pobre; berço de imigrantes rústicos, embora ambiciosos e arrogantes.

Lema: aborrecimento

Mesmo que o futuro possa dizer que a maioria estava errada, é inegável que se dissemina na sociedade norte-americana um **sentimento de saturação e de aborrecimento** com a obsessiva preocupação de parte da comunidade artística em exhibir as chagas de desgraças como a Aids e a miséria.

Quadro 9 - Exemplos de lemas adicionados na anotação semântica com o padrão “sentimento de N”.

Após as etapas iniciais de tratamento da cauda longa, ela ainda continha pouco menos de 700 lemas, número que inviabilizou o processo de análise individual das linhas de concordância por razões de ordem prática para esta pesquisa. Ainda sim, aproximadamente 150 expressões foram distribuídas entre os grupos do Emocionário e a presença de determinados campos lexicais paralelos ao das emoções despontou dos dados.

Durante a análise do “Sentimento de N”, os campos referentes a “nacionalismo”, “espiritualidade” e “parentalidade” já haviam se destacado, como expôs a seção 5.1, e essa característica se reforçou com os lemas pertencentes à cauda longa. Outro campo lexical que também se destacou foi o que denominei “atos automotivados”, com 35 lemas autorreferidos, como “autoflagelação”, “autocobrança” e “autopreservação”. O campo lexical de “nacionalismo”, especialmente, deu lugar a um campo lexical mais abrangente, que poderia ser considerado “coletividade” ou mesmo “pertencimento”, com lemas como “origem”, “enraizamento” e “vizinhança”. A cauda longa também trouxe como antagonista o campo lexical de “estranhamento” ou “não-pertencimento”, com lemas como “desabrigo”, “descolamento” e “diáspora”.

Já no caso dos lemas pertencentes ao campo lexical dos “atos automotivados”, foi conduzida uma análise caso a caso tanto dos lemas encontrados na cauda longa como da distribuição semântica dos lemas equivalentes sem estarem autorreferidos como, por exemplo, o lema da cauda longa “auto-respeito” e seu equivalente não autorreferido “respeito” já possuíam etiquetas dos grupos Humildade e Admirar. O mesmo aconteceu com os lemas “auto-agressividade” e “agrevissidade”, que já eram parte do grupo Fúria; e “autocompaixão”, “compaixão”, “autopiedade” e “piedade” já estavam no grupo Pena. Exceto por estes casos, os demais não possuíam anotação semântica de emoção e foram inseridos em grupos do Emocionário: “autodeterminação”, por exemplo, foi atribuído ao grupo Coragem, do qual já fazia parte o lema “determinação”; “auto-repugnância” e “repugnância” foram inseridas no grupo Ódio; e “auto-reprovação” e “reprovação” no grupo Vergonha. A lista dos lemas que compõem a cauda longa está no Anexo 2, ao final desta tese.

5.2 O mapa emocional do corpo

A ideia de criar um mapa emocional do corpo humano (Ramos et al., 2020) surgiu como uma tentativa de combinar dois projetos da Linguateca – o Esqueleto e o Emocionário – a fim de buscarmos expressões com o corpo humano em português que são usadas para descrever emoção. Além de haver extensas revisões relacionando o uso de palavras do campo semântico do corpo humano com a descrição de emoção (Enfield e Wierzbicka, 2002; Yu, 2002; Sharifian et. al, 2008; Lindquist et. al, 2016; entre outros), os dados do corpus Linguateca (com aproximadamente 40 milhões de tokens na versão 3.2) confirmam essa conexão, mostrando que 49% das frases com uma palavra do corpo também contêm uma palavra de emoção, enquanto 12.9% das frases que contém emoções também incluem uma palavra relacionada ao corpo humano (Ramos et al., 2020:3).

O projeto Esqueleto (Freitas et. al, 2015) estuda o léxico do corpo humano e seus possíveis sentidos através de anotação semiautomática e revisão de grandes corpora. Ele surgiu em vista da percepção da grande relação existente entre expressões de opinião associadas a palavras do corpo humano; além da alta frequência de uso de palavras do corpo humano na anotação semântica de cores na língua portuguesa (Silva e Santos, 2012) e na língua portuguesa de um modo geral.

Na tentativa de evidenciar como os falantes de português se referem ao corpo humano e em quais outros sentidos que não os do corpo propriamente dito, foi feita uma divisão inicial do material entre palavras do corpo humano que fazem referência ao corpo humano e palavras e expressões do corpo humano que se distribuem por campos semânticos adjacentes. Após a primeira análise das ocorrências, foram criadas 15 classes para que as palavras e expressões do corpo humano de outros campos semânticos fossem acomodadas. Dentre elas, as etiquetas [sema="corpo:sentimento"] e [sema="corpo:opinião"], que se diferenciam em definição apenas por estarem com a etiqueta de opinião casos em que o termo marcado apresenta posicionamento ou julgamento explícitos.

As idealizadoras do Esqueleto enfatizam a íntima relação entre sentimento e opinião e o corpo humano, já que algumas emoções provocam alterações fisiológicas; descrever emoções envolve descrever a expressão facial ou corporalmente de seu detentor; e descrever sentimentos ou características de

personalidade com referência a órgãos do corpo humano metonimicamente é uma característica que permeia várias línguas.

Especificamente para o estudo das emoções nesta tese e como outra contribuição à terceira fase do Emocionário, queríamos investigar se as expressões já anotadas para o corpo humano no Esqueleto que fazem referência a emoção (etiquetas [sema="corpo:sentimento"] e [sema="corpo:opinião"]) pertenceriam a um ou mais grupos de emoção descritos no Emocionário. Ou seja, integrar o material anotado do Esqueleto relacionável à descrição de emoção ao léxico do Emocionário.

Assim, partimos de 226 expressões multipalavras (MWEs) anotadas com as etiquetas [sema="corpo:sentimento"] e/ou [sema="corpo:opinião"] no Esqueleto para, em primeiro lugar, atribuímos grupos do Emocionário de acordo com a observação uma a uma, levando em conta os possíveis sentidos emocionais. Desta primeira varredura, foi possível perceber que quanto menos polissêmica e mais lexicalizada é a expressão na língua portuguesa, menos se faz necessário recorrer ao contexto das linhas de concordância. Por exemplo, “gelar a espinha” é uma das expressões que entrou para o grupo Medo sem necessitarmos recorrer ao contexto para confirmação, já que seu sentido é claro para falantes do português. A expressão “ficar de boca aberta”, por sua vez, entrou para os grupos Admirar e Surpresa após a interpretação das linhas de concordância, como mostram os exemplos 74 a 79 a seguir.

Exemplo 74: Então é que o velho ficou com a **boca aberta**, capaz de dar quanto possuía só por tal prenda .

Exemplo 75: Os sábios falam destas coisas e de muitas outras que não sabem, com vozes de papo; e fica toda a gente de **boca aberta** a decorar palavras difíceis que eles inventam e não querem dizer nada .

Exemplo 76: <p>: O pessoal fica de **boca aberta** Com a minha idade fazer aquele crochê .

Exemplo 77: A gente deixava de ir uns poucos de meses num lugar e quando aparecia lá ficava de **boca aberta** vendo tudo mudado: casas novas, negócios sortidos como os da Corte, igreja, circo de cavalinhos, botica e o mato, que é dele?

Exemplo 78: Espantem-se à vontade, podem ficar de **boca aberta**, dar de ombros, tudo; não admito réplica .

Exemplo 79: *par=CCB-A_Filha_do_Arcediago-122*: Acho que a gata lá tem cousa de feitiçaria, porque os homens ficam de **boca aberta** para ela !

Alguns casos que já estavam anotados com a etiqueta “corpo:sentimento”, como “viver de joelhos” e “nervos”, precisaram de anotações de acordo com sentidos de diferentes entradas após a análise das concordâncias. Foram encontradas 44 entradas no AC/DC para “viver de joelhos”. Dessas 44, 26

faziam parte do provérbio popular “melhor morrer de pé do que viver de joelhos” e variantes (como “morrer em pé a viver de joelhos” e outras possíveis combinações). Outras seis ocorrências, nas quais aparecia um mesmo trecho, a expressão “viver de joelhos” foi usada em sua forma literal: “dona Teresinha, analfabeta, 46 anos, cinco filhos, levanta às 4h30 para trabalhar em uma empresa de limpeza, lava escadas de prédios, de segunda a sábado, **vive de joelhos** uma média de dez horas por dia, ganha R\$=70 + R\$=15 do ‘tal’ abono.”. Uma expressão, também com sentido literal, faz uma comparação ao ato religioso de colocar-se de joelhos como representação de devoção: “**Vivi de joelhos** diante de minhas plantas, mais do que eles em suas igrejas.”. As demais ocorrências, ilustradas nos Exemplos 80 a 82, são as que possuem sentido emocional de “submissão” ou “humilhação”²³, parte do grupo Vergonha.

Exemplo 80: Agradeço a oportunidade deste aparte, desejando que em breve o povo brasileiro possa eleger um governo que tenha responsabilidade com a maioria do seu povo, e não um governo submisso, que **vive de joelhos** diante do poder econômico internacional e trata o povo da maneira como o Governo Fernando Henrique vem tratando .

Exemplo 81: Se pagarem a dívida não terão condições de dar comida ao seu povo, de obter crescimento econômico, pois **vivem de joelhos** perante o capital financeiro volátil, que entra neste País e sai dele na hora em que bem entende .

Exemplo 82: Afinal, por que comemorar a independência de um país que **vivia de joelhos** diante de uma cruel e sanguinária ditadura ?

Logo, a alteração cabível a essa expressão foi criar uma outra etiqueta específica para o provérbio, com o lema “morrer em pé a viver de joelhos” e possíveis variações, e incluir na EVP “viver de joelhos” o sema do grupo Vergonha.

A palavra “nervos” estava anotada com a etiqueta “corpo:sentimento_faculdade” e contava com quase 10 mil ocorrências. Ao longo da leitura da amostra aleatória apresentada pelo AC/DC, foi possível destacar algumas características para serem confirmadas com apoio de outras ferramentas do projeto: o uso literal de “nervos” quando sucedido por um adjetivo e o uso emocional quando antecedido por um pronome possessivo ou em algumas expressões. Para confirmar o uso literal, foi conduzida uma busca a fim de especificar quais adjetivos estão conectados a “nervos” e, das mais de 2 mil ocorrências, a listagem evidencia que as referências são à faculdade do corpo humano, como mostra a Figura 6, com um recorte dos adjetivos que sucedem a

²³ definição corroborada pelo Dicionário Brasileiro de Fraseologia de José Pereira da Silva (2013: 780): “**De joelhos**. Com os joelhos em terra; em atitude de submissão, de derrotado, humilhado (ANS). “Quero ver se você é capaz de me pedir perdão de joelhos, na frente de seus homens.” (GCC)”

palavra “nervos”. A resolução para as regras foi retirar a etiqueta de sentimento de “nervos” quando aparece combinada com um adjetivo.

Procura: [word="nervos"] @[pos="ADJ"]
Distribuição de lema
Corpo: os corpos todos v. 8.1

2203 casos.

Distribuição

Houve **294** valores diferentes de lema.

periférico	466	sensorial	27
craniano	428	facial	26
óptico	86	femoral	26
mediano	43	tibial	26
espinhal	41	motor	24
intercostal	38	sensitivo	21
isquiático	33	lombar	20
pudendo	27	simpático	20

Figura 6 - Distribuição das combinações de adjetivos para o lema “nervos”.

Já nas ocorrências de “nervos” precedida por um pronome possessivo foi possível detectar vários sentidos emocionais distintos nos contextos das entradas. E, por mais que a palavra “nervos” isoladamente não tenha um sentido emocional específico, ela claramente está sendo usada para descrever emoções como excitação, nervosismo, inquietação e animação, como mostram os Exemplos 83 a 88. A resolução para essa anotação foi adicionar uma regra com o pronome possessivo antecedendo a palavra “nervos” na qual ela ganhe a etiqueta do grupo Genérico, que marca a presença de uma emoção não específica.

Exemplo 83: Aproveitando um momento de mais calma para os **meus nervos**, disse-lhe para mudar de conversa :

Exemplo 84: Quando os poisava na miudinha figura de morenita que eu era então, toda a minha carne se arrepiava numa tremura e os **meus nervos** vibravam desagradavelmente .

Exemplo 85: No plebiscito inflamado de todos os **seus nervos** o amor venceia .

Exemplo 86: E em seguida, às noites, no salão, -- aonde sempre acudia uma ou outra família vizinha, -- era ainda ela o prestigioso centro das atenções, o fulcro espiritual do diálogo, neste morno e sossegado ambiente, sentindo-se na posse plena dos **seus nervos** e em todo o facetado fulgor do seu espírito .

Exemplo 87: Há alguma coisa mais forte do que o sentimento, acredita: é a vida que nos corre nas veias, que vibra em todos os **nossos nervos**, que fremente em relâmpagos sensuais à flor da carne .

Exemplo 88: Os poetas não têm versos, têm cavaquinhos, violões e a voz para dobrar e quebrar os **nossos nervos** .

Outra adequação nas regras precisou acontecer para acomodar as expressões de várias palavras envolvendo “nervos”: “guerra de nervos”; “nervos

à flor da pele”; “crise de nervos”; “ataque de nervos”; “sofrer dos nervos”; e “dar nos nervos”. A regra de anotação da EVP “guerra nos nervos” compreendia apenas os casos em que “guerra” era escrita com a letra “g” maiúscula, então precisou ser alterada para passar a considerar casos com a palavra “guerra” com “g” minúsculo. Às expressões “nervos à flor da pele”, “crise de nervos” e “ataque de nervos” foram adicionados o sema do grupo Medo, que também compreende palavras como “nervoso” e “inquieto”. A expressão “sofrer dos nervos” ganhou a etiqueta do grupo Desespero, que compreende sentidos de sofrimento de um modo geral, enquanto a expressão “dar nos nervos” foi incluída no grupo Fúria, que compreende sentidos equivalentes a “enervar” e “irritar”.

6

Revisão e reorganização do léxico: limpeza dos grupos de emoção

No ano de 2020, foi realizada uma revisão em sete dos 24 grupos que compõem o léxico (“Desespero”, “Esperança”, “Humildade”, “Pena”, “Satisfação”, “Saudade” e “Surpresa”). Um dos principais objetivos à época era mensurar o desafio de analisar aspectos semânticos de uma língua e tomar consciência das características da análise de emoção, alternando leituras distantes e aproximadas, para aprimorar a anotação semântica do léxico e auxiliar a criação de subsídios para a anotação semi-automática do campo semântico das emoções na língua. Boa parte da análise dos lemas da revisão está nas discussões e nos exemplos do capítulo 2 desta tese, onde foram evidenciados características e desafios da análise de emoção na língua portuguesa.

A revisão dos grupos envolveu recursos associados aos estudos linguísticos com corpus como linhas de concordância, listas de frequência, distribuição por campo semântico e lemas; alternou análises de dados nas formas quantitativa e qualitativa através de leitura distante e leitura aproximada; e contou com consulta em dicionários e ferramentas de pesquisa na internet. Na presente seção, descrevo a metodologia desenvolvida a partir dessa revisão de forma detalhada, fazendo uso de exemplos retirados do próprio corpus no processo de aplicação do método para ilustrar os passos. A metodologia segue um total de sete passos que podem ser agrupados em três momentos: triagem, busca por informações adicionais e redefinição dos grupos.

O primeiro momento tem como objetivos verificar e confirmar as anotações prévias dos lemas e estudar linhas de concordância para buscar regras para os lemas com múltiplos sentidos (emocionais e não-emocionais). Este estudo pretende verificar a precisão e abrangência dos possíveis sentidos dos lemas analisados, considerando os contextos nos quais estão sendo usados. Esse momento inicial envolve os passos 1, 2 e 3 da metodologia e está descrito em detalhes na seção 6.1, a seguir.

O segundo momento compreende os passos 4 e 5 da metodologia e está detalhado na seção 6.2. Ele busca por informações que porventura validem o

sentido emocional dos lemas fora do corpus. Ou seja, é uma tentativa de mapear outros usos possíveis dos lemas em análise, através de pesquisas em dicionários, ou mesmo mostrar outros usos emocionais mais atuais do lema pelos falantes, caso haja, a partir de ferramentas de busca on-line, por exemplo,

No terceiro momento, os passos 6 e 7 têm como finalidade reconfigurar os grupos do léxico do Emocionário (conforme os critérios elencados no capítulo 4 desta tese) e concluem o processo de revisão. Os grupos revisados tomam forma a partir da identificação de novos lemas que podem ser adicionados ao Emocionário; lemas que precisam de regras específicas que separem sentidos emocionais de não-emocionais; e da manutenção, exclusão ou realocação de lemas nos grupos de palavras de emoção. Este é o processo final da metodologia e está explicado e exemplificado na seção 6.3 deste capítulo.

6.1

Ferramentas de anotação: corte-e-costura e Vislcg3

Nesta seção, descrevo e exemplifico duas ferramentas usadas no processo de anotação no AC/DC: o corte-e-costura e o VISLCG3. O projeto VISL (*Visual Interactive Syntax Learning*) foi idealizado por Eckhard Bick a partir do final da década de 1990. Consiste em um sistema eletrônico que visa ao processamento morfossintático de uma gama de línguas baseado em um analisador automático (*parser*) que utiliza a Gramática Constritiva (*Constraint Grammar – CG*) como base, e é composto por milhares de regras capazes de analisar morfológica e sintaticamente qualquer texto nele inserido. O VISL trabalha juntamente com um conjunto de orações controladas para sistematizar fenômenos sintáticos.

A versão original analisa os campos de forma linear, indicando a quantos tokens à direita ou à esquerda elas aparecem a partir da posição zero da frase, geralmente ocupada pelo sujeito/sintagma nominal. A descrição gramatical automática é baseada na CG, mas pode ser diversificada para sistemas com outros usos específicos, como anotações com códigos de cores, texto anotado ou árvores sintáticas (*treebanks*). Os *treebanks* do VISL se constituem a partir da adição de dados de gramáticas de frases estruturais (*PSG-grammars*) aos corpora anotados automaticamente pela CG e são considerados a parte central

do projeto. Já o VISL3 (Bick 2009; 2015) é uma terceira e mais atual versão do compilador de regras e estruturas do âmbito da Constraint Grammar cujo foco foi introduzir inovações para melhorar o desempenho do sistema, também facilitando o manejo do usuário. É um sistema de código aberto, que pode ser utilizado com diversos fins.

O PALAVRAS (Bick, 2000) é o *parser* que analisa a língua portuguesa no AC/DC. Efetivamente, foi a partir dele que o sistema VISL foi desenvolvido. Com a etiquetagem de palavras no âmbito da CG, a anotação acontece de forma horizontal e baseada em palavras, isto é, a anotação de cada palavra refere-se ao seu núcleo sintático na frase, a relação imediata de dependência. Graças ao PALAVRAS, também foi possível implementar ao sistema a análise automática de orações subordinadas e um sistema de marcadores de dependência, que viabilizaram a transformação da anotação horizontal em *treebanks*.

Mesmo com etiquetas nos níveis semântico, sintático e morfológico, ele alcança um nível de precisão geral de 98% (Bick, 2000:1). Essa etiquetagem diversificada possibilita buscas por classes de palavras, função sintática e semântica de lemas ou palavras específicas. A anotação de emoção é escrito em VISL3 pela equipe da Linguateca após o PALAVRAS e é válida para todos os corpora do AC/DC. As regras completas do VISL3 para emoção são parte do Emocionário e estão disponíveis no site da Linguateca²⁴.

As regras, que seguem o formalismo da CG, são divididas entre restritivas, para os níveis morfológico e semântico, e produtivas e restritivas, para o nível sintático. Como explica Bick (2000):

A Gramática Constritiva tenta formalizar o processo cognitivo de desambigüização em um conjunto de regras que constroem – por intermédio de condições contextuais – qual das interpretações possíveis para cada palavra será escolhida ou rejeitada. Essas regras compilam-se num *parser* "reducional" que - nos níveis de morfologia e semântica - seleciona a etiqueta certa. No nível sintático o *parser* contém regras tanto produtivas como restritivas, essas mapeando etiquetas ambíguas de função sintática, aquelas rejeitando ou selecionando etiquetas através do contexto. (BICK, 2000a:2)

Para esta tese, o interesse recai principalmente sob as regras produtivas, utilizadas para as etiquetas semânticas e, portanto, as palavras que descrevem emoções. Abaixo, os Exemplos 89 a 91 demonstram como as regras produtivas podem rejeitar ou selecionar as etiquetas de acordo com o contexto dos corpora.

Exemplo 89: ADD (%EMO:SURPRESA) TARGET ("choque") (-1 ("em"));

Exemplo 90: # regras mais finas para a palavra crença

²⁴ Lista de regras do VISL3 comuns a todos os corpora disponíveis em: https://www.linguateca.pt/aceso/corpos/obras/regras_emodizer.utf8.txt.

ADD (%EMO:ESPERANCA) TARGET ("crença") (1 ("em")) (2 ("que")) (3* (SUBJ));
 ADD (%EMO:ESPERANCA) TARGET ("crença") (1 ("inabalável"));
 Exemplo 91: ADD (%EMO:HUMILDADE) TARGET RESPEITO (NOT -1 ("a"))
 (NOT 1 ("de")) (NOT -1 ("dizer"));

A regra do Exemplo 89 se refere ao grupo Surpresa e serve para ativar a etiqueta semântica de surpresa para as combinações nos contextos em que o lema choque é antecedido pela preposição “em”. O Exemplo 90 mostra duas regras para o lema crença, pertencente ao grupo Esperança. A primeira indica que “crença”, se sucedido por oração que se inicie com a combinação “em que” seguida de um verbo no subjuntivo, deve ser anotada como Esperança. A segunda anota os casos em que o lema crença aparece imediatamente seguido do adjetivo “inabalável”. Já o Exemplo 91 ilustra três regras que são parte do grupo Humildade e rejeitam a anotação de emoção para o lema “respeito” nas combinações “respeito de”, “a respeito” e “respeito” antecedido pelo verbo “dizer” em qualquer conjugação. Em todas as regras, o nome “TARGET” sempre faz referência à palavra que seleciona ou rejeita a etiqueta. No caso das regras que selecionam a etiqueta, as partículas que se ligam diretamente ao “TARGET” estão demarcadas de acordo com sua posição em relação ao “TARGET”, sendo números positivos se aparecem após e negativos quando aparecem antes dele. As regras que rejeitam, como é o caso do Exemplo 91, funcionam da mesma forma sendo apenas acrescentada a regra “NOT”, indicando a rejeição daquela etiqueta semântica para o contexto.

O programa corte-e-costura, por sua vez, é parte do projeto AC/DC. Foi idealizado por Mota e Santos (2010) e surgiu a partir da necessidade de criar regras ou modificar as já existentes na anotação anterior do AC/DC, mas de uma forma prática e que evitasse a necessidade de aprender novas linguagens de programação por parte do anotador. Para isso, o corte-e-costura foi desenhado de forma que o linguista que saiba manusear o formato das expressões de busca usado na interface de usuário do AC/DC (IMS-CWB) tenha também autonomia para trabalhar na criação e modificação das regras em textos já altamente processados. Por consequência, o corte-e-costura também é útil para dar conta dos erros da anotação principal do AC/DC, que é feita de forma totalmente automática.

Os principais usos do corte-e-costura são na anotação semântica: a adição dos atributos “sema” e “grupo” e identificação de expressões multpalavras (EVPs ou *MWEs*). Dentre os objetivos da ferramenta estão a criação de mecanismos e procedimentos para anotação semântica semiautomática (Santos

& Mota, 2010:1441) que conseguissem resultados positivos de forma rápida e que fossem reaplicáveis a cada vez que o AC/DC passasse por mudanças na anotação.

O corte-e-costura é composto por quatro conjuntos de regras (Santos & Mota, 2010:1442): positivas, negativas, de especialização e recursivas. Em relação ao campo semântico das emoções, as duas primeiras são as mais relevantes. Um corpus de teste foi criado para testar algumas regras confeccionadas para o corte-e-costura sugeridas ao longo da exploração do AC/DC a serem adicionadas ao Emocionário. Desta forma, os testes seriam feitos sem que a estrutura final do AC/DC fosse prejudicada, caso alguma anotação estivesse incorreta. Ou seja, era adicionado ao corpus de teste um conjunto de frases contendo exemplos das expressões cujas anotações semânticas seriam mudadas. Depois que a regra era aplicada e funcionava, aquelas frases eram trocadas por outras, para que um novo conjunto de regras fosse testado.

As regras positivas anotam um lema ou MWE que carrega o valor semântico da etiqueta em questão – o interesse desta tese recai sobre o domínio da emoção. Testamos regras positivas com os lemas “falta”, “querer”, “inchar” e “cara de caso”. Considerando os sentidos analisados nas frases em que aparecem no AC/DC, os lemas que aparecem no Quadro 10, juntamente com as respectivas regras, sofreram alterações de anotação. “Falta” foi anotado no grupo Saudade no Emocionário nos contextos “fazer falta”, “sentir a falta” e “sentir a minha/sua/nossa/vossa falta”. “Querer” foi adicionado ao grupo Amor quando aparece como característica atribuída a uma pessoa. “Inchar”, por sua vez, foi adicionado ao grupo Orgulho quando coocorre com as palavras “adulação”, “valimento”, “conto”, “consagração” e “distinção”. Já a MWE “cara de caso” passou por três mudanças: foi anotada como uma MWE; o lema “cara” foi anotado como “corpo:sentimento” no contexto dessa expressão; e a MWE foi adicionada ao grupo Surpresa.

falta

```
[lema="fazer"] a:[word="falta"] >> a:[sema="emo:saudade"]
[lema="sentir"] [word="a"] a:[word="falta"] >> a:[sema="emo:saudade"]
[lema="sentir"] [word="a"] [pos="DET_poss.*"] a:[word="falta"] >> [sema="emo:saudade"]
```

querer

```
[word="querid[oa]s*" & pos="N|ADJ.*"] >> [sema="emo:amor"]
[word="querid[oa]s*" & pos="V.*" & temcagr="PCP"] >> [sema="emo:amor"]
```

inchar


```
a:[lema="inchar"] [word="com"] [pos="DET.*"]
[lema="adulação|valimento|conto|consagração|distinção"] >> a:[sema="emo:orgulho"]
```

cara de caso

```
a:[lema="cara"] b:[lema="de"] c:[lema="caso"] >> <mwe
sema="corpo:sentimento_emo:surpresa" lema="cara=de=caso" pos="N"> a: b: c:
</mwe> a:[sema="corpo:sentimento"]
```

Quadro 10 - Exemplos de regras positivas adicionadas ao Corte-e-Costura.

As regras negativas, no contexto das palavras de emoção, servem para os casos nos quais um lema é do campo semântico de emoção na maioria das vezes em que aparece no corpus mas, em alguns contextos específicos, é parte de outro campo semântico. Então, as regras negativas são criadas para que esses lemas percam a anotação semântica de emoção.

O Quadro 11 ilustra as regras negativas feitas para o lema “apreensão”, que é parte do grupo Medo. Após a adaptação das regras através do corte-e-costura, “apreensão” passa a não ter anotação semântica quando coocorre as palavras “arma”, “droga”, “bem”, “documento”, “armamento” e “polícia”, e quando aparece nos padrões “apreensão de grande/pequena quantidade de” e “mandato/mandado de busca e apreensão”.

apreensão

```
[word="apreensão"] >> [sema="0"]
a:[lema="apreensão"] [lema="de(\+o)*"]
[lema="arma|droga|bem|documento|armamento"] >> a:[sema="0"]
[lema="mandado|mandato"] [word="de"] a:[lema="apreensão"] >> a:[sema="0"]
[lema="busca"] [word="e"] a:[lema="apreensão"] >> a:[sema="0"]
a:[lema="apreensão"] [lema="policial"] >> a:[sema="0"] # para a Literateca
a:[lema="apreensão"] [word="e"] [lema="tortura"] >> a:[sema="0"]
a:[lema="apreensão"] [word="pela"] [lema="polícia|Polícia.*"] >> a:[sema="0"]
a:[lema="apreensão"] [word="de"] [pos="NUM.*"] >> a:[sema="0"]
a:[lema="apreensão"] [word="de"] [lema="grande|pequeno"] [lema="quantidade"] >>
a:[sema="0"]
```

Quadro 11 - Exemplos de regras negativas adicionadas ao Corte-e-Costura.

As regras de especialização e recursivas são menos relevantes para a notação das emoções, mas também merecem uma explicação breve: as regras de especialização²⁵ consistem na adição de regras para contextos nos quais os lemas ou expressões são transportados para um campo semântico mais específico que o que está em foco. Essa especialização não ocorreu com as palavras de emoção pois as anotações do Emocionário já vêm do VISLCG3 anotadas com a etiqueta de emoção e o grupo, mas um tipo de especialização foi feita com a anotação de sentimento dentro do campo semântico de corpo

²⁵ Para exemplos mais específicos, ver trabalhos com Cor (Silva et al., 2008 e também em <https://www.linguateca.pt/acesso/Arcolris.pdf>) e com o Corpo Humano (Freitas et al., 2015 e também em <https://www.linguateca.pt/acesso/Esqueleto/>)

humano. O Exemplo 92 mostra a expressão “cabeça erguida”, que já era anotada semanticamente no Esqueleto (seção 5.2 desta tese) como “sema=corpo:sentimento_posicao”. Essa anotação, após o processo de combinação das expressões do corpo humano com os grupos do Emocionário, passou a ser “sema=corpo:sentimento_posicao_emo:orgulho”.

Exemplo 92: par=AA-Casa_de_Pensão-2206: Logo que lhe falaram nisso, pediu secamente a nota de suas despesas, pagou-a, e retirou-se muito calmo, assoviando, de mão no bolso, **cabeça erguida**, na mesma fleuma inalterável com que costumava sair todas as manhãs para o trabalho.

Por último, as regras recursivas são usadas sobretudo em relações de coordenação para desambiguação de elementos não pertencentes ao campo semântico sob análise, e são aplicadas após as outras três regras. Assim como as regras de especialização, não são muito relevantes para a anotação das emoções, e não é possível exemplificá-las com este campo semântico, já que nenhum trabalho foi feito com esse tipo de regra durante esta pesquisa. Para exemplos mais específicos, é possível consultar o manual do corte-e-costura²⁶, disponível publicamente no site da Linguateca.

6.2

Primeiro momento: Triagem dos lemas – passos 1, 2 e 3

No primeiro momento, seleciono o lema para análise e busco pelas etiquetas semânticas que ele carrega (pela opção “distribuição por anotação semântica” no AC/DC), concretizando o Passo 1 da metodologia. No exemplo da Figura 7 está o lema “satisfeito”, que carrega as etiquetas dos grupos Felicidade e Satisfação²⁷. No caso de “satisfeito”, portanto, após os resultados da busca, apenas confirmei nas linhas de concordância se os usos estão marcados corretamente e segui adiante com a análise. Como os exemplos estavam de acordo com a anotação, não houve aplicação dos passos 2 a 5 para o lema “satisfeito”.

²⁶ Disponível em <https://www.linguateca.pt/acesso/corte-e-costura/>.

²⁷ Além da etiqueta “pred:emo”, que faz referência aos predicadores humanos e é fruto da pesquisa de mestrado conduzida por Silva (2021) no corpus OBRas. Para efeitos da revisão do grupo “satisfação”, o relevante nesta busca é que todas as entradas de “satisfeito” estão anotadas como “emo:satisfeito”.

Procurar: [lema="satisfeito"] OK

Resultado:

- Concordância
- Distribuição das formas (*word*)
- Distribuição dos lemas (*lema*)
- Distribuição da categoria gramatical (PoS) (*pos*)
- Distribuição do tempo verbal e/ou do caso pronominal (*temcagr*)
- Distribuição de pessoa e/ou número (*persnum*)
- Distribuição do género morfológico (*gen*)
- Distribuição da função sintáctica (*func*)
- Distribuição por variante do português (*variante*)
- Distribuição por género de texto (*classe*)
- Distribuição por campo semântico (*sema*)
- Distribuição por grupo (de cor, roupa, etc.) (*grupo*)
- Distribuição por corpo (*corpo*)

Procura: [lema="satisfeito"]
 Distribuição de **sema**
 Corpo: os corpos todos v. 7.6

1938 casos.

Distribuição

Houve 2 valores diferentes de **sema**.

emo:feliz_emo:satisfeito 1871
 emo:feliz_emo:satisfeito_pred:emo 67

Figura 7 - Distribuição por campo semântico do lema “satisfeito”.

O exemplo da Figura 7 também mostra que as etiquetas podem se somar, por isso, ao fazer a busca para saber se um lema específico está anotado com alguma emoção, se adicionados os símbolos “.*” antes e depois dos caracteres “emo” na expressão de busca, a possibilidade de encontrar resultados é ampliada.

No passo 2, aqui ilustrado pelo lema “pedido”, busco pelas linhas de concordância das entradas que estão anotadas como emoção (*sema*="emo"), como mostra a expressão de busca²⁸ na Figura 8, a fim de confirmar se a anotação condiz com os contextos de uso do lema “pedido”. A partir desses exemplos, fica claro que ele não tem sentido emocional em particular.

Concordância

Procura: [lema="pedido" & sema=".*emo.*"].

Apresenta-se uma amostra aleatória de 5000 das 31921 ocorrências encontradas.

par=FSP940918-048-580: Para fazer o **pedido** é preciso ir a uma agência da CEF .

par=FSP950901-080-1012: Para as prefeituras, atende em média 50 **pedidos** por ano, por meio de projetos, sobretudo nas áreas de infra-estrutura urbana e rural e de habitação .

par=71943: Ao decidir o **pedido**, 1] o juiz admitirá como verdadeiros os fatos 2] que, por meio do documento ou da coisa, a parte pretendia provar :

par=ext490284-soc-92a-1: O Ministério recebeu a missiva, mas nada respondeu até agora, tal como nada dissera em relação aos anteriores **pedidos** de reunião com o secretário de Estado, segundo se queixa a Associação .

par=FSP941002-129-2179: Os **pedidos**, que atingiram 27 mil, segundo a Fiat, serão feitos por computadores ligados diretamente com a fábrica .

par=73423: «Anula-se o acórdão que não contém dispositivo preciso acerca do provimento parcial que deu à apelação, desconhecendo-se se algum dos **pedidos** formulados pelos autores mereceu ou não acolhimento» (RSTJ 43/310) .

Figura 8 - Mostragem de linhas de concordância do lema “pedido” anotado como palavra de emoção.

²⁸ A expressão de busca correspondente é [lema="pedido" & sema=".*emo.*"]

A partir das linhas de concordância da Figura 8, é possível observar que não há diferenciação nos sentidos independentemente da anotação, que foi inicialmente feita de forma automática²⁹. No entanto, ao longo da observação das regras já existentes no projeto das emoções, foi interessante perceber que há uma regra que exclui corretamente da anotação de emoção a expressão “a pedido de”. Ainda assim, a leitura das linhas de concordância dos dois casos (com e sem anotação semântica) mostrou que o sentido do lema “pedido” não é emocional em nenhum dos contextos. Ilustrativamente, a partir de uma busca³⁰ no corpus pelos complementos de “pedido”, as dez combinações mais frequentes foram “pedido de demissão”; “pedido de liminar”; “pedido de cassação”; “pedido de autorização”; “pedido de prisão”; “pedido de desculpa”; “pedido de impeachment”; “pedido de urgência”; “pedido de extradição”; e “pedido de falência”. Dessas, a única combinação que indica emoção é “pedido de desculpa”, que já faz parte do grupo “Humildade” por meio de uma regra de anotação específica do Emocionário³¹.

No passo 3, busco por exemplos de linhas de concordância das entradas que estão anotadas com outros campos semânticos ou estão sem anotação semântica (`sema!=".emo.*"`) para confirmar os usos do lema em análise nos contextos do AC/DC. Para a leitura das linhas de concordância, nos casos em que há muitas ocorrências, a ferramenta apresenta uma listagem randômica e, a partir da amostragem automática, seleciono as linhas para leitura, também aleatoriamente.

De um total de 39.601 entradas do lema “choque”, mais de 90% das entradas estão sem anotação semântica de emoção. A leitura das linhas de concordância, juntamente com a lista de regras já inseridas no sistema, evidenciam algumas combinações que ganham sentido em outros campos semânticos, como mostram os Exemplos 93 a 96, e não carregam sentido emocional.

Exemplo 93: *par=ext1444674-soc-96a-2*: João sofreu um **choque** eléctrico que lhe causou a morte imediata .

Exemplo 94: <p>: O caminhão do **choque** deixou o local minutos depois .

Exemplo 95: *par=FSP950821-030-352*: O blindado levava munição e explosivos, que teriam sido detonados com o **choque** .

²⁹ Como explicado na seção 2.3 desta tese, sobre as primeiras fases do Emocionário

³⁰ A expressão de busca correspondente é `[lema="pedido" & sema="*.emo.*"] [lema="de"] @[pos="N"]`

³¹ A regra mencionada está no VISLCG3 como “ADD (%EMO:HUMILDADE) TARGET ("pedido") (1 ("de")) (2 ("desculpa"))”

Exemplo 96: <p>: Informações preliminares indicavam que a causa do naufrágio foi o **choque** do barco com um tronco de árvore.

6.3

Segundo momento: Informações adicionais – passos 4 e 5

Os Passos 4 e 5 confirmam ou buscam por informações adicionais sobre os sentidos dos lemas analisados fora dos corpora, a fim de validar os usos emocionais também lemas fora do corpus. Acontecem a partir de pesquisas em dicionários, ferramentas de fora da linguística com corpus, mas que são referência em léxico, e pesquisas on-line, que possibilita encontrar usos mais recentes do lema pelos falantes que, porventura, ainda não tenham sido adicionados a dicionários. Esses passos só precisam ser aplicados em casos nos quais não seja possível confirmar o uso do sentido emocional do grupo apenas pelas linhas de concordância, ou em casos nos quais o lema carrega sentidos diversos, como é o caso de “arrebatar”, que foi usado para ilustrar a aplicação desses passos.

De acordo com os resultados da distribuição semântica de “arrebatar”, as etiquetas atribuídas ao lema foram dos grupos Desespero e Fúria, como atesta a Figura 9. Ressalto que o grupo Fúria não foi revisado nesta fragmento da metodologia. Logo, o interesse aqui recai sobre as anotações correspondentes a Desespero.

Procura: [lema="arrebatar"]
Distribuição de sema
Corpo: os corpos todos v. 7.6

3027 casos.

Distribuição

Houve 2 valores diferentes de sema.

emo:furia 1690
emo:desespero_emo:furia1337

Figura 9 - Distribuição semântica do lema “arrebatar” no AC/DC.

Os Exemplos 97 a 101 listam algumas linhas de concordância analisadas durante a aplicação do Passo 2 e demonstram como esses lemas que carregam

anotação semântica de Desespero não transmitem nenhum dos possíveis sentidos para o grupo³².

Exemplo 97: *par=147*: Mas havia também livros que falavam de fraquezas como as minhas e um deles foi como uma vaga que, durante vários meses, me **arreatou** .

Exemplo 98: *par=149*: Entretanto, passei a dar o meu lugar no autocarro às pessoas de idade, oferecia-me para as ajudar a carregar os sacos quando saíam do supermercado, nunca tentava pendurar-me na parte traseira dos carros, não partia nada, não atirava aos pássaros com uma fisga, olhava para o chão ao andar procurando não pisar as formigas ou os escaravelhos e, até mesmo quando na Primavera colhia flores para os meus pais, na companhia de Geir ou de outros, sofria pensando nas vidas que desse modo estava a **arreatar** .

Exemplo 99: *par=18*: Como este apelido de Cubas lhe cheirasse excessivamente a tanoaria, alegava meu pai, bisneto do Damião, que o dito apelido fora dado a um cavaleiro, herói nas jornadas da África, em prêmio da façanha que praticou **arreatando** trezentas cubas ao mouros .

Exemplo 100: *par=95*: Esteiros foi a obra que mais me **arreatou** nos anos sessenta, seguido de Engrenagem, do mesmo escritor, que faziam parte da biblioteca subversiva do meu pai .

Exemplo 101: *par=183*: Um dinamarquês traçoeiro conseguiu **arreatar** consigo a bola para o centro, deu uma reviravolta e precipitou-se na minha direcção .

O Passo 4 consiste, então, em procurar pela definição no dicionário para confirmar o uso com sentido de emoção equivalente ao do lema dentro dos contextos das linhas de concordância e/ou ao grupo no qual ele estava inserido antes da revisão.

Ao buscar pela definição de “arreatar” nos Dicionários Aulete Digital e Priberam (Figura 10), pude confirmar sua utilização com sentido emocional, porém não compatível com os sentidos contidos no grupo Desespero especificamente.

arreatar

(ar.re.ba.tar)

v.

1. Tirar com força ou violência; ARRANCAR [td. : *Arreatou o bilhete incriminatório*] [tdr. + a, de : *A policia arreatou dos cambistas todos os ingressos*.]
2. Carregar consigo (o vento) ao soprar [td. : *Súbita ventania arreatou as folhas secas (do chão)*.]
3. Extasiar(se), encantar(-se). [td. : *A música clássica o arreata*] [tr. + com : *Arreatou -se com a cena final*]
4. Tornar(-se) colérico; ENFURECER(-SE) [td. : *Nada o arreata mais do que a falsidade*] [tr. + com : *Arreatou -se com a injúria*]
5. Suscitar, provocar; ARRANCAR [td. : *O filme arreatou muitos elogios*] [tdr. + a, de : *Seu dempenho arreatou á'da plateia merecida ovação*]
6. Conquistar. [td. : *O tenista arreatou o primeiro lugar*]

[F.: a² - + rebat(e) + -ar².]

ar·re·ba·tar - Conjugar

(a- + rebate + -ar)

verbo transitivo

1. Tirar para si com violência.
2. Privar de, roubar.
3. Levár à força e com violência.
4. Arrancar (alguma coisa) com violência (das mãos de outrem).
5. Induzir, levar.
6. Entusiasmar.
7. Extasiar.

verbo pronominal

8. Encolerizar-se.

³² Explicação do grupo Desespero: “O grupo “Desespero” engloba sentidos como estado de espírito ou sofrimento daquele que passa por inúmeras dificuldades e aflições e não tem como superá-las ou acredita que não o possa fazer; desesperança; sofrimento moral extremo, misto de aflição, angústia, descontrolo e tormento.” – Disponível em: <https://www.linguateca.pt/Gramateca/Emocionario.html>

Figura 10 - Definição de “arrebatar” nos Dicionários Aulete Digital e Priberam, respectivamente.

Assim, nem dentro dos contextos das linhas de concordância, nem isoladamente, o lema “arrebatar” carrega sentidos equivalentes ao do grupo Desespero. A vantagem de se pesquisar em dicionários é ter acesso a informação de usos daquela palavra em contextos que porventura não se apresentem na busca em corpus.

O Passo 5 é a tentativa de encontrar usos semelhantes do lema sob análise e assim ratificar o uso daquele lema com sentido emocional pelo falante do português nas ferramentas de pesquisa on-line. O Passo 5 só acontece quando o dicionário não apresenta nenhuma definição emocional isolada, mas o contexto das linhas de concordância claramente indica emoção. No caso do lema “arrebatar”, como a definição do dicionário não levava a uma aproximação ao sentido do grupo Desespero, busquei no Google (Figura 11) – mas poderia ter sido em outra ferramenta, e por isso o título diz “ferramenta de busca” – e tampouco consegui confirmar o uso emocional que justificasse sua manutenção no grupo Desespero.

No entanto, fui redirecionada ao lema “arrebato”, cuja definição do dicionário na primeira resposta da busca é “sentimento de...” e já fornece uma pista de outro lema que pode vir a ser adicionado ao léxico do Emocionário, absorvido ao processo do Passo 6, explicado na seção 6.4 a seguir.

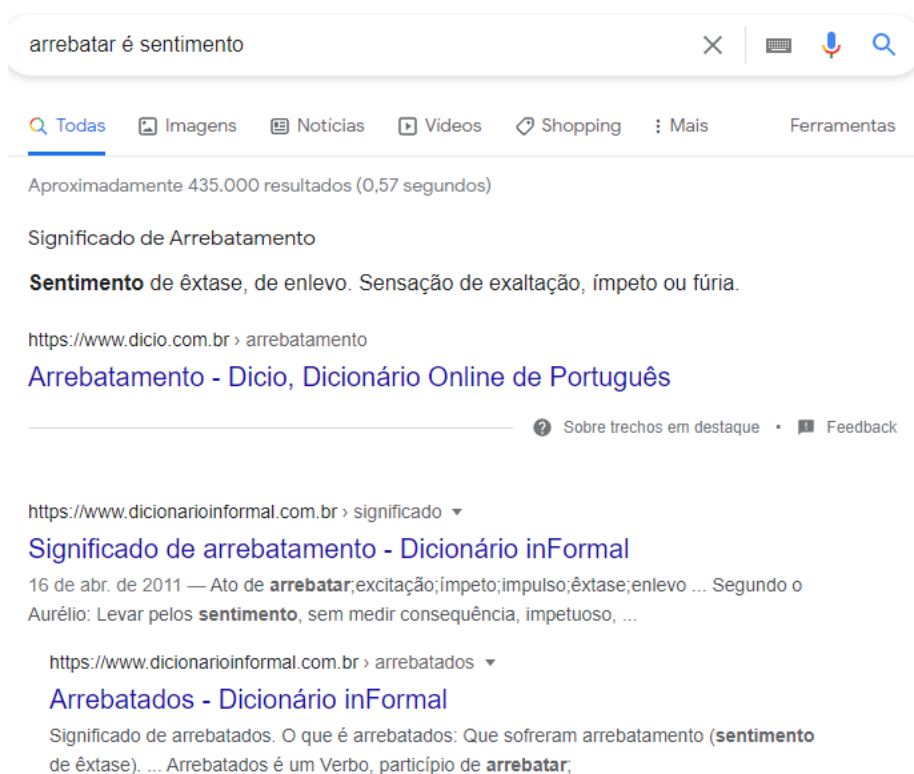


Figura 11 - Resultado de busca por “arrebatar é sentimento” no Google.

Vale ressaltar que o Passo 5 só é aplicado na metodologia como último recurso ao qual, na maior parte das vezes, não foi necessário recorrer ao longo da revisão dos sete grupos listados no início desta seção. Quando utilizado, como no caso de “arrebatar”, a pesquisa foi feita na ferramenta on-line com frases “X é sentimento” e/ou “X é emoção”.

6.4

Terceiro momento: Redefinição dos grupos – passos 6 e 7

Trata-se do último momento da metodologia e une os Passos 6 e 7. O Passo 6 revisa os achados durante a leitura das linhas de concordância nos Passos 2 e 3 e as pesquisas em outras ferramentas nos Passos 4 e 5 de modo a detectar coocorrências de outras palavras de emoção que possam ser inseridas neste ou em um dos outros grupos do Emocionário, caso apareçam em diversos exemplos. O grupo Pena, por exemplo, ganhou os lemas “clemência”, “clemente”, “compassivo” e “condolência”. Como ilustram os Exemplos 102 a 107, os lemas “clemência” e “compassivo” coocorriam com os lemas em análise – neste caso, o lema “compaixão”, assim como apareciam em comum em

definições de dicionários, e foram selecionados para passarem pela mesma metodologia que os lemas que já estavam nos grupos. Após serem analisados individualmente respeitando os mesmos passos da metodologia, foram adicionados ao grupo Pena.

Exemplo 102: <p>: As grandes tradições morais que se fundam na **compaixão** são a cristã e a budista ainda que possam ser encontrados elementos compassivos no hinduísmo, no islamismo e no judaísmo 60 .

Exemplo 103: <p>: Lembra que a justiça só pode ser compreendida em conjunto com a equidade e a imparcialidade e que as diferenças que o compassivo reforça dificilmente podem produzir relações fundadas na equidade, até pela simples razão de que a **compaixão** não tem por objetivo transformar as condições materiais que conduzem à desigualdade (Caponi, 2000) .

Exemplo 104: <p>: Assim o nosso chefe, que procura Mostrar-se compassivo, não descansa Com estas poucas obras: passa a dar-nos de sua **compaixão** maiores provas .

Exemplo 105: <p>: Que impulsos de **compaixão**, de clemência, não moveram o augusto coração de um bom rei, quando pôs os olhos na calamidade pública ?

Exemplo 106: <p>: Não espere otimismo, clemência ou **compaixão** do mestre polonês .

Exemplo 107: <p>: Espera-se assim diminuir a responsabilidade do acusado, produzindo **compaixão** e clemência .

Finalmente, no Passo 7, mantenho ou retiro o lema previamente listado no grupo após as análises individuais. Algumas das resoluções foram a exclusão dos lemas “pedido” e “pedir” do grupo Humildade; e a manutenção com criação de regras para incluir no léxico do Emocionário apenas os casos em que os lemas “esperar”, do grupo Esperança, e “alegre” e “jubilação”, do grupo Satisfação, carregam o sentido emocional. Essas regras podem ser consultadas no Anexo 4 desta tese, onde as mudanças que ocorreram nos grupos revisados do léxico estão documentadas especificamente.

7 Resultados: o Emocionário hoje

Considerando todo o trabalho realizado no Emocionário ao longo das análises e explorações em corpora da presente tese, é possível identificarmos dois principais pontos que resultaram em mudanças no léxico de palavras de emoção que é entregue ao usuário do AC/DC. As mudanças da seção 7.1 foram de ampliação dos grupos de emoção e aconteceram a partir de explorações do padrão “sentimento de N” e das expressões com o corpo humano anotadas pelo Esqueleto, ambas em corpora do AC/DC, detalhadas no capítulo 5 desta tese. A seção 7.2 retrata o segundo momento das mudanças, resultado da revisão completa de sete grupos do léxico do Emocionário (detalhada no capítulo 6), que modificou as listagens finais de lemas de todos os grupos revisados. Ao final do capítulo é feito um apanhado de todos os resultados e seus impactos no produto final do léxico do Emocionário.

7.1 Resultados da ampliação e validação de padrões no Emocionário

A partir de buscas pelo padrão “sentimento de N” e por expressões com palavras do corpo humano já anotadas semanticamente como sentimento ou opinião pelo Esqueleto, foi possível ampliar as entradas do Emocionário assim como validar seu conteúdo. Dentre os 843 lemas com ocorrência maior que três entradas, advindos da busca do “sentimento de N” em todo o AC/DC, 742 foram atribuídos a grupos de emoção do léxico do Emocionário. Já os lemas da cauda longa, com menos de três ocorrências no corpus, aproximadamente 150 foram inseridos em grupos do léxico até o momento e pouco mais de 500 não foram analisados até a conclusão desta pesquisa. O padrão se mostrou confiável, pois 90% das palavras encontradas eram de sentimento.

Das 226 expressões combinando sentidos de emoção e opinião provenientes das anotações do Esqueleto com palavras do corpo humano, as 131 anotadas com a etiqueta [sema=“corpo:sentimento”] foram atribuídas a um

ou mais grupos do Emocionário; dezenove expressões de um total de 64 com a etiqueta [sema="corpo:opinioao"] entraram em pelo menos um dos grupos; e das 24 que carregam etiquetas de sentimento e opinião, sendo [sema="corpo:sentimento_opinioao"], dezesseis foram adicionadas a grupos do Emocionário. Dentre as sete expressões anotadas com [sema="corpo:sentimento"] e mais um subgrupo do Esqueleto, sendo estes "doença", "faculdade", "outros" e "posição"³³, todas elas foram incluídas em pelo menos um dos grupos, dentre Coragem, Desespero, Genérica, Medo, Orgulho e Satisfação.

O resultado da exploração das etiquetas de sentimento e opinião do Esqueleto confirma que o padrão de anotação é confiável da perspectiva do estudo das emoções, sendo em torno de setenta por cento das expressões conciliáveis com os grupos do Emocionário; evidencia que a descrição de emoção está presente também nos casos de opinião, ainda que em menor porcentagem – o que já era de certa forma esperado, dado que no português não necessariamente há emoção anexada a toda opinião – pouco menos de trinta por cento dos casos anotados somente como "opinião" receberam uma etiqueta de um grupo do Emocionário, ao passo que aproximadamente 65 por cento das etiquetas de [sema="emo:sentimento_opinioao"] foram adicionadas a pelo menos um grupo do Emocionário; enriquecendo o léxico de emoções com 154 novas expressões que combinam os campos semânticos do corpo humano e da emoção. Em relação à atribuição de grupos, o grupo de emoções Genéricas foi o que mais acumulou novas palavras e expressões, com 22 no total, seguido de 21 novas entradas para o grupo Desespero e 15 novas entradas para os grupos Insatisfação e Supresa. O único grupo ao qual não foi atribuída nenhuma nova entrada foi o grupo Saudade, mas em uma busca pela expressão "[lema="coração"] [lema="estar"] [pos="ADJ"]" – na qual "coração" carrega a etiqueta [sema="corpo:sentimento_emo:gen"] – coocorrendo com o lema "saudade", vemos que há sim menção de saudade com o corpo humano, como mostram os exemplos 108 e 109.

Exemplo 108: <p>: É Jean quem começa a conversa, perguntando a Dhomini se ele estava com mais saudades de Sabrina, sua namorada do programa, ou de Manoela, sua namorada de há mais tempo: Jean: . O **coração tá foda, né, cara?** ... **De quem você tá sentindo mais saudade ?**

Exemplo 109: ! ... 165 O desejo de regressar à pátria é também uma preocupação deste poeta cujo **coração está repleto de saudades** .

³³ [sema="corpo:sentimento_doenca"]; [sema="corpo:sentimento_faculdade"]; [sema="corpo:sentimento_outros"]; e [sema="corpo:sentimento_posicao"].

O Anexo 3 desta tese compreende a configuração das expressões do Esqueleto divididas por grupos do Emocionário.

7.2

Resultados da revisão do léxico do Emocionário

Sete dos 24 grupos do Emocionário foram integralmente revistos através da metodologia desenvolvida especificamente para esta análise, relatada no capítulo 6 desta tese: Desespero, Esperança, Humildade, Pena, Satisfação, Saudade e Surpresa.

Em termos quantitativos, a Tabela 4 elenca os números envolvidos na revisão, separados por “lemas antes da revisão”; “lemas excluídos”; “lemas adicionados”; “lemas adicionados a outros grupos”; e “lemas após a revisão”. O Grupo Desespero foi o que teve o maior número de mudanças, com uma média de 20% dos lemas de antes da revisão excluídos e substituídos por novos lemas. Nos grupos Esperança, Humildade, Saudade e Surpresa houve modificações menores que 10% em relação a exclusões e adições de lemas; o grupo Pena teve cerca de 30% de aumento de lemas; e o grupo Satisfação teve em torno de 15% de aumento do número de lemas em relação a antes da revisão.

Como resultado da revisão dos sete grupos, então, os números não parecem tão distintos do cenário de antes da revisão. No entanto, essa revisão parcial do léxico tornou possível mensurar a dificuldade de analisar aspectos semânticos dentro da língua com foco na descrição de emoção. A etapa inicial da revisão também permitiu delinear características e desafios da análise de emoção em português, detalhadas no capítulo 4 desta tese; aprimorar uma parcela da anotação do Emocionário; e documentar uma metodologia que pode ser replicada futuramente nos grupos restantes ou em outras tarefas de revisão com objetivos semelhantes.

Grupo	Lemas antes da revisão	Lemas excluídos	Lemas adicionados	Lemas adicionados a outros grupos	Lemas após a revisão
Desespero	40: afligir, aflito, agonia, agonizante, agonizar, alvoroçado, alvoroçar, alvoroço, amargura, amargurado, amargar, angustiar, arrebatado, arrebatarse, desalentador, desalento, desanimador, desanimar, desânimo, desconfiança, desconfiar, desconsolação, desconsolado, desconsolar, desconsolo, desesperação, desesperadamente, desesperado, desesperador, desesperança, desesperançar, desesperante, desesperar, desesperativo, desespero, insinuar, mortificado, mortificar, suplicante, suplicar.	7: alvoroçado, alvoroçar, alvoroço, arrebatado, arrebatarse, desesperativo, insinuar.	8: angústia, angustiado, atormentado, atormentar, desconfiado, desesperançoso, súplica, tormento.	11: amargura, atormentado, atormentar, tormento, desconfiado, desconfiança, desconfiar, mortificar súplica, suplicante, suplicar.	41: afligir, aflito, agonia, agonizante, agonizar, amargura, amargurado, amargar, angústia, angustiado, angustiar, atormentado, atormentar, desalentador, desalento, desanimador, desanimar, desânimo, desconfiado, desconfiança, desconfiar, desconsolação, desconsolado, desconsolar, desconsolo, desesperação, desesperadamente, desesperado, desesperador, desesperança, desesperançar, desesperançoso, desesperante, desesperar, desespero, mortificado, mortificar, súplica, suplicante, suplicar, tormento.
Esperança	36: acreditar, afiançar, almejar, ansioso, augurar, auspicioso, confiante, confiança, confiança, confiar, confidenciar, crença, desejar, desejo, desejoso, entrever, esperança, esperançadamente, esperançar, esperançosamente, esperançoso, esperar,	2: crença, expectativa.	2: crer, promissoramente.	2: entrever, augurar.	36: acreditar, afiançar, almejar, ansioso, augurar, auspicioso, confiante, confiança, confiar, confidenciar, crer, desejar, desejo, desejoso, entrever, esperança, esperançadamente, esperançar, esperançosamente, esperançoso, esperar, expectante, expectativa,

Grupo	Lemas antes da revisão	Lemas excluídos	Lemas adicionados	Lemas adicionados a outros grupos	Lemas após a revisão
	expectante, expectativa, expectativa, expectante, expectativa, expectação, fé, optimismo, optimista, otimista, oxalá, prometedor, prometedoramente, promissor.				expectação, expectante, expectação, fé, optimismo, optimista, otimista, oxalá, prometedor, prometedoramente, promissor, promissoramente.
Humildade	24: abnegação, apelar, confessar, deferência, despretensiosamente, despretensioso, despretensão, humildade, humilde, humildemente, humildosamente, humilde, modestamente, modesto, modéstia, pedido, pedir, pedir desculpa, reconhecer, respeitador, respeito, respeitosamente, respeitoso.	2: pedido, pedir.	0	0	22: abnegação, apelar, confessar, deferência, despretensiosamente, despretensioso, despretensão, humildade, humilde, humildemente, humildosamente, humilde, modestamente, modesto, modéstia, pedir desculpa, reconhecer, respeitador, respeito, respeitosamente, respeitoso.
Pena	12: apiedar, apiedar-se, compaixão, desculpar, dó, misericordioso, misericórdia, piedade, piedoso, solidariedade, solidarizar, solidário.	0	4: clemência, clemente, condolência, compassivo	0	16: apiedar, apiedar-se, clemência, clemente, condolência, compaixão, compassivo, desculpar, dó, misericordioso, misericórdia, piedade, piedoso, solidariedade, solidarizar, solidário.
Satisfação	37: Agradado, agradar, agrado, agradável, alegre, alegria, aprazimento, aprazível, comprazer, comprazimento, contentamento, contente,	1: agradado	6: apreciar, contentar, deleitar, deleitar-se, deleite, júbilo	1: apreciar	42: agradar, agrado, agradável, alegre, alegria, aprazimento, aprazível, apreciar, comprazer, comprazimento, contentamento, contentar,

Grupo	Lemas antes da revisão	Lemas excluídos	Lemas adicionados	Lemas adicionados a outros grupos	Lemas após a revisão
	desfrutar, desfrute, distrair, divertimento, divertir, entreter, gosto, gozar, gozo, gratificante, jubilação, lisonjeiramente, lisonjeiro, prazer, regalado, regalo, regozijo, repleto, saciado, saciar, saciação, saciedade, satisfatório, satisfação, satisfeito.				contente, deleitar, deleitar-se, leite, desfrutar, desfrute, distrair, divertimento, divertir, entreter, gosto, gozar, gozo, gratificante, jubilação, júbilo, lisonjeiramente, lisonjeiro, prazer, regalado, regalo, regozijo, repleto, saciado, saciar, saciação, saciedade, satisfatório, satisfação, satisfeito.
Saudade	10: banzo, falta, lembrança, nostalgia, nostálgico, recordação, saudade, saudosamente, saudosista, saudoso .	0	1: saudosismo.	0	11: banzo, falta, lembrança, nostalgia, nostálgico, recordação, saudade, saudosamente, saudosista, saudoso, saudosismo.
Surpresa	41: admirado, admirar, admirar-se, admiração, aparvalhado, assombrar-se, atônito, atónito, banzar, banzo, boquiaberto, chocar, choque, embasbacado, embasbacar, espantado, espantar, espanto, estupefacto, estupefato, imprevisto, imprevisível, impromptu, inesperadamente, inesperado, pasmado, pasmar, pasmo, perplexidade, perplexo, repentino, rompante, sobressaltar, sobressalto, surpreendente, surpreendentemente,	3: chocar, choque, súbito.	2: maravilhado, maravilhar.	0	40: admirado, admirar, admirar-se, admiração, aparvalhado, assombrar-se, atônito, atónito, banzar, banzo, boquiaberto, embasbacado, embasbacar, espantado, espantar, espanto, estupefacto, estupefato, imprevisto, imprevisível, impromptu, inesperadamente, inesperado, maravilhado, maravilhar, pasmado, pasmar, pasmo, perplexidade, perplexo, sobressaltar, sobressalto, surpreendente, surpreendentemente,

Grupo	Lemas antes da revisão	Lemas excluídos	Lemas adicionados	Lemas adicionados a outros grupos	Lemas após a revisão
	surpreender, surpreendido, surpresa, surpreso, súbito.				surpreender, surpreendido, surpresa, surpreso.

Tabela 4 - Panorama quantitativo da revisão do léxico do Emocionário

8 Considerações finais

A parte inicial dessa pesquisa buscou contextualizar os estudos relacionados a descrição de emoção ao longo do tempo, inclusive em áreas paralelas aos Estudos da Linguagem, com o intuito de registrar que se trata de um tema que é objeto de interesse de diferentes áreas de pesquisa e considerado complexo e muito extenso. Por não haver consenso na literatura, esta tese refuta a ideia de emoções básicas inerentes a todas as culturas, também porque a cultura é viva e mutável, o que inevitavelmente impacta no léxico e nas mudanças linguísticas. O que pode ser considerado básico para uma realidade pode não existir em outra, como exemplificado na seção 1.1 com a passagem da antropóloga Laura Bohannan por uma família de Ifaluks que não conseguia conceber conceitos decorrentes de experiências baseadas em vivências e conjuntos de regras ocidentais (Rezende e Coelho, 2010).

O terceiro capítulo descreve léxicos que existem para a língua portuguesa na área de Análise de Sentimento e explica que o léxico do Emocionário não se insere na mesma classificação, pois sua motivação está em identificar e categorizar emoção. Se o interesse está em avaliar textos, o léxico do Emocionário é uma excelente maneira de olharmos para a língua, já que é formado a partir de copora anotados e revisados, um ponto chave para que toda a análise possa ser conduzida.

O processo da construção da metodologia para revisão do léxico do Emocionário se deu através de uma análise do próprio léxico à luz de três questionamentos: quais caminhos utilizar para confirmar se uma palavra é de emoção ou não; em que casos a palavra pode ser considerada uma descrição de emoção; e em que grupo (ou grupos) de emoção cada palavra se encaixa. Essa análise evidenciou características da análise de descrição de emoção, como a relevância de expressões lexicalizadas como “rua da amargura” e “fundo do poço”; tipos de argumentos dos verbos que podem causar variação no sentido da frase; a utilização de frases ou expressões que manifestam negação; a utilização de prefixos e sufixos, bem como variações de tempo e modo verbal,

que acarretam em acréscimos semânticos e mudança de significado; e as cargas semânticas emocionais que as palavras carregam, como “pedir”, que não tem sentido emocional isoladamente, e “suplicar” e “implorar”, que compartilham do mesmo significado de “pedir” porém com peso emocional. O quarto capítulo também trouxe à tona desafios na análise de emoção, como palavras e expressões polissêmicas; palavras homógrafas pertencentes a diferentes classes gramaticais; e a dificuldade de detectar diferentes sentidos em palavras que compartilham a mesma classe gramatical de forma automática tendo como base apenas a informação morfossintática. A análise ressalta a relevância do contexto onde os objetos da análise estão inseridos e da anotação em corpus, capaz de minimizar ou mesmo solucionar muitos dos desafios aqui encontrados.

O sétimo capítulo agrupa os resultados de todas as explorações e análises da pesquisa descritas nos capítulos quatro a seis e funciona também como uma demonstração de que o objetivo geral desta tese, de criar subsídios para a anotação do campo semântico das emoções na língua portuguesa, foi alcançado. Apesar de os objetivos de “descrever o campo semântico das palavras usadas para indicar emoção e sentimento em grandes corpora de língua portuguesa” e “apresentar critérios para atribuir emocionalidade a uma palavra” terem sido considerados objetivos específicos, acredito agora que apenas o tempo de uma pesquisa de doutorado, ainda que seja um processo longo e denso, não seria suficiente para cumpri-los em sua totalidade. Em parte, pela grande dificuldade atrelada à tarefa de descrever emoções propriamente dita, mas também por se tratar de um processo contínuo, já que a língua, bem como a forma que os falantes se utilizam de seus aspectos, é mútavel.

A partir da busca por padrões da língua, mais especificamente as pesquisas pelo padrão “sentimento de N” e pelas expressões anotadas pelo Esqueleto nos forneceram pistas de outros grupos de emoção além dos previamente existentes no Emocionário: “Ausência” e “Outra”; possibilitaram a detecção de campos lexicais pouco mencionados na literatura sobre emoção, como “coletividade”, “estranhamento”, “espiritualidade”, “parentesco” e “atos automotivados”. Agrupando todas as explorações e análises em corpora, foram adicionados aproximadamente novecentos novos lemas e expressões ao léxico do Emocionário.

O desenvolvimento da metodologia para revisão de um terço dos grupos do léxico do Emocionário (capítulo 6) mostra que os objetivos de “validar os grupos de emoção do léxico do Emocionário para classificar as palavras que descrevem emoção” e de léxico “reavaliar, revisar e reorganizar o léxico e

anotações semânticas do Emocionário” foram apenas parcialmente alcançados. No entanto, essa revisão, ainda que parcial, estruturou uma metodologia que pode ser replicada futuramente, tanto em outros grupos do léxico do Emocionário, quanto em outras tarefas de revisão com objetivos semelhantes e até mesmo em outras línguas, levando em conta adaptações necessárias. Assim, acredito ser coerente afirmar que esse processo de fato contribuiu para a consolidação do AC/DC como um material de referência na área de anotação de emoções e para a constituição de material mais refinado e com um léxico mais abrangente.

A respeito de trabalhos futuros que dialogariam diretamente com a pesquisa aqui apresentada, aponto como uma possível nova exploração o aprofundamento da anotação de emoção dos personagens de literatura lusófona, a partir da anotação de predicadores humanos já existente no AC/DC; uma investigação sobre hierarquias entre emoções, tendo como motivação a observação dos grupos Fúria e Ódio; e a validação de casos difíceis de serem categorizados por outros falantes de português através de análises consensuais.

9 Referências bibliográficas

ALLEN, B.; BLEVINS, C.; HEUSER, R. **The emotions of London**. Disponível em: <<https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>>. Acesso em: 26 mar. 2023.

ARAÚJO, N. Vista de longe, a literatura é o que desaparece... (Acerca de um fracasso programático em Franco Moretti). Em: ARAÚJO, N.; SOARES, M. V.; WERKEMA, A. (Eds.). **Variações sobre o romance**. Rio de Janeiro: Makunaima, 2016. p. 259–272.

BALAGE FILHO, P. P.; PARDO, T. A. S.; ALUISIO, S. M. **An evaluation of the Brazilian Portuguese LIWC Dictionary for sentiment analysis**. Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology – STIL. 2013.

BICK, E. **The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework**. [s.l.] University of Aarhus, 2000a.

BICK, E. **Ensino de Sintaxe Portuguesa na Internet**. II Encontro Internacional de Português-Língua Estrangeira. Universidade de São Paulo, 2000b.

BICK, E. PALAVRAS - A Constraint Grammar-Based Parsing System for Portuguese. Em: SARDINHA, T. B.; DE LOURDES SÃO BENTO FERREIRA, T. (Eds.). **Working with Portuguese Corpora**. [s.l.] Bloomsbury Academic, 2014. p. 279–302.

BICK, E.; DIDRIKSEN, T. **CG-3 - Beyond Classical Constraint Grammar**. 20th Nordic Conference of Computational Linguistics (NODALIDA). Linköping University Electronic Press, 2015.

BODDICE, R. **The history of emotions**. [s.l.] Manchester University Press, 2015.

CARVALHO, P. et al. **Liars and saviors in a sentiment annotated corpus of comments to political debates**. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, OR, USA: Association for Computational Linguistics, 2011.

CARVALHO, P.; SILVA, M. J. Sentilex-pt: principais características e potencialidades. Em: SIMÕES et al. (Eds.). **Linguística, Informática e Tradução: Mundos que se Cruzam**. [s.l.] Oslo Studies in Language 7(1), 2015. p. 425–438.

CHEN, Y.; SKIENA, S. **Building sentiment lexicons for all major languages**. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg, PA, USA: Association for Computational Linguistics, 2014.

DIAS DA SILVA, B. C. et al. **O TeP: construção de um thesaurus eletrônico para o português do Brasil**. Programa e Resumos do II Congresso Internacional da ABRALIN. 2001.

EKMAN, P. Universals and cultural differences in facial expressions of emotion. **Nebraska Symposium on Motivation**, v. 19, p. 207–283, 1971.

EKMAN, P. Basic Emotions. Em: DALGLEISH, T.; POWERS, M. J. (Eds.). **Handbook of Cognition and Emotion**. Sussex, Inglaterra: John Wiley & Sons, 1999. p. 45–60.

ELLIS, J. M. **Language, thought, and logic**. Evanston, IL, USA: Northwestern University Press, 1993.

ENFIELD, N. J.; WIERZBICKA, A. Introduction: The body in description of emotion. **Pragmatics & cognition**, v. 10, n. 1–2, p. 1–25, 2002.

ENGELMANN, A. A natureza do problema investigado. Em: **Os estados subjetivos: uma tentativa de classificação de seus relatos verbais**. São Paulo: Ática, 1978. p. 15–82.

FREITAS, C. et al. Sparkling Vampire... lol! Annotating Opinions in a Book Review Corpus. Em: ALUÍSIO, S.; TAGNIN, S. E. O. (Eds.). **New Language Technologies and Linguistic Research: A Two-Way Road**. Newcastle, Inglaterra: Cambridge Scholars, 2014. p. 128–146.

FREITAS, C. et al. O léxico do corpo e anotação de sentidos em grandes corpora: o projeto Esqueleto. **Revista de Estudos da Linguagem**, v. 23, n. 3, p. 641, 2015.

GRAYSON, S. et al. **Discovering Structure in Social Networks of 19th Century Fiction**. 8th ACM Conference on Web Science. 2016.

HEARST, M. A. **Automatic acquisition of hyponyms from large text corpora**. Proceedings of the 14th conference on Computational linguistics. Morristown, NJ, USA: Association for Computational Linguistics, 1992.

KLINGER, R.; SAMAT, S.; REITER, N. **Automatic Emotion Detection for Quantitative Literary Studies: A case study based on Franz Kafka's "Das Schloss" und "Amerika"**. Digital Humanities 2016: Conference Abstracts. Cracóvia, Polônia: 2016.

LE BRETON, D. **Experiências da dor: uma antropologia**. , 13 mar. 2019.

LEITÃO DE ALMEIDA, M. L. et al. A hipótese de corporificação da categorização e do léxico. Em: LEITÃO DE ALMEIDA, M. L. et al. (Eds.). **Linguística Cognitiva em foco: morfologia e semântica do português**. Rio de Janeiro: Publit, 2010. p. 187–204.

LINDQUIST, K.; GENDRON, M.; SATPUTE, A. Language and Emotion: Putting Words into Feelings and Feelings into Words. Em: BARRETT, L. F.; LEWIS, M.; HAVILAND-JONES, J. M. (Eds.). **Hanbook of Emotions (4th Edition)**. New York, NY, USA: The Guildford Press, 2016. p. 579–594.

MAIA, B.; SANTOS, D. Language, emotion, and the emotions: The multidisciplinary and linguistic background: Language, Emotion, and the Emotions! **Language and linguistics compass**, v. 12, n. 6, p. e12280, 2018.

MOHAMMAD, S. M.; TURNEY, P. D. Crowdsourcing a word-emotion association lexicon. **Computational intelligence**, v. 29, n. 3, p. 436–465, 2013.

MORETTI, F. **A Literatura Vista De Longe**. [s.l.] Arquipélago, 2008.

MORETTI, F. **Distant Reading**. Munique, Germany: Verso, 2013.

MOTA, C. **Anotação de emoções nos corpos do AC/DC**. Disponível em: <<https://www.linguateca.pt/documentos/Mota2013.pdf>>.

MOTA, C.; SANTOS, D. Emotions in natural language: a broad-coverage perspective. 2015.

MOTA, C.; SANTOS, D. **Corte e costura no AC/DC: auxiliando a melhoria da anotação nos corpos**. Disponível em: <<http://www.linguateca.pt/aceso/corte-e-costura.pdf>>.

MOTA, C.; SANTOS, D. **Emotions in natural language: a broad-coverage perspective**. Disponível em: <<https://www.linguateca.pt/aceso/EmotionsBC.pdf>>.

NABI, J. **Uma abordagem inteligente para a saúde mental**. Disponível em: <<https://www.nexojornal.com.br/externo/2019/05/25/Uma-abordagem-inteligente-para-a-saude-mental>>. Acesso em: 27 mar. 2023.

ORTONY, A.; CLORE, G. L.; COLLINS, A. **The cognitive structure of emotions**. Cambridge, England: Cambridge University Press, 2011.

PASQUALOTTI, P. R.; VIEIRA, R. WordnetAffectBR: uma base lexical de palavras de emoções para a língua portuguesa. **RENOTE**, v. 6, n. 1, 2008.

PENNEBAKER, J. W. et al. Linguistic inquiry and word count: Liwc 2001. **Mahwah, NJ: Lawrence Erlbaum Associates**. 2001.

PENNEBAKER, J. W. et al. The development and psychometric properties of LIWC2007. **Austin, TX, LIWC. Net**, 2007.

PLUTCHIK, R. **The emotions: facts, theories, and a new model**. New York, NY, USA Random House, , 1962.

PLUTCHIK, R. The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. **American scientist**, v. 89, n. 4, p. 344–350, 2001.

RAMOS, B. **Leitura Distante e categorização de sentimentos representados pelo corpo humano na Literatura**. Primeiro Encontro de Leitura Distante em Português. Oslo, Noruega: Universidade de Oslo, 2019.

RAMOS, B. C. **Descrição de uma metodologia desenvolvida para revisão de um léxico de palavras de emoção**. XIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL 2021). Sociedade Brasileira de Computação, 2021. Disponível em: <<http://dx.doi.org/10.5753/stil.2021.17819>>

RAMOS, B.; FREITAS, C. **“Sentimento de quê?”: uma lista de sentimentos para a Análise de Sentimentos**. STIL - Symposium in Information and Human Language Technology. Salvador, BA: 2019.

RAMOS, B.; SANTOS, D.; FREITAS, C. **Looking at body expressions to enrich emotion clusters**. (M. J. B. Finatto et al., Eds.) Proceedings of the Digital Humanities and Natural Language Processing Workshop at the 14th International Conference on the Computational Processing of Portuguese Language. 2020.

REZENDE, C. B.; COELHO, M. C. **Antropologia das Emoções**. Rio de Janeiro: Editora FGV, 2010.

SANTOS, A. G. L.; BECKER, K.; MOREIRA, V. **Um estudo de caso de mineração de emoções em textos multilíngues**. III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM). 2014.

SANTOS, D. **Providing access to language resources through the World Wide Web: the Oslo Corpus of Bosnian Texts**. (A. Rubio et al., Eds.) First International Conference on Language Resources and Evaluation. Granada, Espanha: 1998.

SANTOS, D. Um centro de recursos para o processamento computacional do português. **DataGramZero - Revista de Ciência da Informação**, v. 3.1, 2002.

SANTOS, D. et al. **Linguateca: um Centro de Recursos Distribuído para o Processamento Computacional da Língua Portuguesa**. (G. De Ita Luna, O. F. Chávez, M. O. Galindo, Eds.) International workshop "Taller de Herramientas y Recursos Lingüísticos para el Español y el Portugués", IX Iberoamerican Conference on Artificial Intelligence (IBERAMIA). Puebla, México: 2004.

SANTOS, D. Corporizando algumas questões. Em: Tagnin, S.; Vale, O. A. (Eds.) **Avanços da Lingüística de Corpus no Brasil**. USP, São Paulo: Editora Humanitas, 2008. p. 41-66.

SANTOS, D. A Gramateca, os seus objetivos e alguns exemplos. **PUC Rio**, 2014.

SANTOS, D. Literature studies in Literateca: between digital humanities and corpus linguistics. Em: DOERR, M. et al. (Eds.). **Humanists and the digital toolbox: In honour of Christian-Emil Smith Ore**. Oslo, Noruega: Novus forlag, 2019. p. 89–109.

SANTOS, D. et al. Leitura distante em português: resumo do Primeiro Encontro. **Matlit Revista do Programa de Doutorado em Materialidades da Literatura**, v. 8, n. 1, p. 279–298, 2020.

SANTOS, D.; BICK, E. **Providing Internet access to Portuguese corpora: the AC/DC project**. (M. Gavrilidou et al., Eds.) Proceedings of the Second International Conference on Language Resources and Evaluation (LREC). 2000.

SANTOS, D.; COSTA, L. A Linguateca e o projecto “Processamento Computacional do português”. **Terminómetro - Número especial nº 7 - A terminologia em Portugal e nos países de língua portuguesa em África**, p. 63–69, 2005.

SANTOS, D.; FREITAS, C.; BICK, E. OBras: a fully annotated and partially human-revised corpus of Brazilian literary works in public domain. **OpenCor**, v. Canela, RGS, Brasil, 2018.

SANTOS, D.; FREITAS, C.; LOPES, J. M. **Ler e estudar a literatura lusófona como parte da literatura mundial: recursos para leitura distante em português**. HD Rio. 2018.

SANTOS, D.; MAIA, B. Language, emotion, and the emotions: A computational introduction. **Language and linguistics compass**, v. 12, n. 6, 2018.

SANTOS, D.; MOTA, C. **Experiments in Human-computer Cooperation for the Semantic Annotation of Portuguese Corpora**. Seventh International Conference on Language Resources and Evaluation (LREC). European Language Resources Association (ELRA), 2010.

SANTOS, D.; MOTA, C. A admiração à luz dos corpos. Em: SIMÕES, A. et al. (Eds.). **Linguística, Informática e Tradução: Mundos que se Cruzam. Homenagem a Belinda Maia**. Oslo, Noruega: OSLa, 2015. v. 7p. 57–77.

SANTOS, D.; SIMÕES, A.; MOTA, C. Broad coverage emotion annotation. **Language resources and evaluation**, v. 56, n. 3, p. 857–879, 2022.

SHARIFIAN, F. et al. Culture and language: Looking for the “mind” inside the body. Em: SHARIFIAN, F. et al. (Eds.). **Culture, Body, and Language: Conceptualizations of Internal Body Organs across Cultures and Languages**. Berlin, NY, USA: Mouton de Gruyter, 2008. p. 3–26.

SIMÕES, A.; SANTOS, D. **Nos bastidores da Gramateca: uma série de serviços**. (L. A. Alemany et al., Eds.)| Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish. 2014.

SOUZA, M. et al. **Construction of a Portuguese Opinion Lexicon from multiple resources**. Brazilian Symposium in Information and Human Language Technology. 2012.

TABOADA, M. Sentiment analysis: An overview from linguistics. **Annual review of linguistics**, v. 2, n. 1, p. 325–347, 2016.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. **Journal of language and social psychology**, v. 29, n. 1, p. 24–54, 2010.

TOMKINS, S. S. **Affect imagery consciousness: Volume I - The Positive Affects**. [s.l.] Springer Publishing Company, 1962.

VALE, O. A. As opiniões nas expressões e a expressão da opinião. Em: LAPORTE, E.; SMARSARO, A.; VALE, O. (Eds.). **Dialogar é preciso**:

Linguística para processamento de linguas. [s.l.] PPGEL/UFES, 2013.
p. 259–267.

VIEIRA, R.; SOUZA, M. **Construction of a Portuguese Opinion Lexicon from multiple resources.** Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana. 2011.

WIEBE, J.; WILSON, T.; CARDIE, C. Annotating expressions of opinions and emotions in language. **Language Resources and Evaluation**, v. 39, n. 2–3, p. 165–210, 2005.

WIERZBICKA, A. **Studies in emotion and social interaction: Emotions across languages and cultures: Diversity and universals: Diversity and universals.** Cambridge, England: Cambridge University Press, 1999.

Anexo 1

O Anexo 1 lista os lemas divididos nos Grupos 1 e 2 pela pesquisa “sentimento de N”, que agrupam sentimentos convencionais e não-convencionais, respectivamente. Os lemas que estão ao final de cada lista já faziam parte do Emocionário antes da pesquisa.

GRUPO 1

acolhimento	completude	desolação
afinidade	compreensão	despeito
aflição	comprometimento	despreparo
agressão	comunhão	diferença
alteridade	conformidade	doação
altruísmo	conformismo	dor
ameaça	conforto	dúvida
amorosidade	confusão	eficácia
apatia	constrangimento	egoísmo
arrependimento	cordialidade	empatia
asco	crise	encanto
autoconfiança	crítica	entrega
autodefesa	cuidado	esgotamento
autodesvalorização	culpa	estabilidade
auto-eficácia	culpabilização	estranhamento
autonomia	cumplicidade	estresse
benevolência	curiosidade	exaustão
bondade	deleite	excitação
cansaço	depressão	familiaridade
caos	desamparo	fidelidade
carência	descaso	força
catástrofe	descrença	fraqueza
cautela	desencantamento	fraternidade
certeza	desencanto	frustração
companheirismo	desgaste	grandeza
competição	desigualdade	harmonia

honestidade	maldade	responsabilidade
hospitalidade	miséria	ressentimento
igualdade	nacionalismo	revanche
incerteza	negação	ruptura
incômodo	negatividade	sarcasmo
incompletude	nojo	saudosismo
inconformidade	normalidade	segurança
inconformismo	otimismo	sensibilidade
indignidade	perda	sobrevivência
inevitabilidade	perplexidade	solidão
inquietação	pessimismo	sucesso
inquietação	potência	sufocamento
insegurança	preocupação	tédio
insignificância	prepotência	tensão
instabilidade	rebeldia	tolerância
insucesso	reciprocidade	tragédia
intimidação	rejeição	tranquilidade /
intolerância	remorso	tranqüilidade
intranquilidade /	reprovação	trunfo
intranqüilidade	repúdio	vulnerabilidade
invalidez	repugnância	xenofobia
irresponsabilidade	resignação	

adoração emo:amor	cólera emo:fúria
afecto / afeto emo:amor	comoção emo:gen
afetividade emo:amor	consternação emo:infeliz
agonia emo:desespero	decadência emo:vergonha
alívio emo:alívio	dedicação emo:amor
amargura emo:desespero	desagrado emo:insatisfeito
animosidade emo:ódio	desalento emo:desespero
ansiedade emo:medo	desamor emo:ódio
apego emo:amor	desconfiança emo:desespero
assombro emo:medo	descontentamento emo:infeliz
atração emo:amor_emo:desejo	desejo emo:desejo_emo:esperança
auto-confiança emo:esperança	desesperança emo:desespero
choque emo:suspresa	desespero emo:desespero
cobardia emo:medo	desgraça emo:infeliz

desprazer emo:insatisfeito
 desprezo emo:vergonha
 desrespeito emo:orgulho
 devoção emo:amor
 dó emo:pena
 embaraço emo:vergonha
 entusiasmo emo:feliz
 espanto emo:surpresa
 esperança emo:esperança
 euforia emo:feliz
 expectativa emo:esperança
 falta emo:saudade
 felicidade emo:feliz
 glória amo:feliz
 impaciência emo:fúria
 infelicidade emo:infeliz
 ira emo:fúria_emo:ódio
 irritação emo:fúria
 júbilo emo:feliz
 luto emo:infeliz
 mágoa emo:infeliz
 nostalgia emo:saudade
 ódio emo:ódio
 otimismo emo:esperança
 paixão emo:amor

pânico emo:medo
 pavor emo:medo
 esperança emo:esperança
 raiva emo:fúria_emo:ódio
 receio emo:medo
 regozijo emo:feliz_emo:satisfeito
 saudade emo:saudade
 simpatia emo:amor
 sofrimento emo:infeliz
 solidariedade emo:pena
 submissão emo:vergonha
 surpresa emo:surpresa
 ternura emo:amor
 tristeza emo:infeliz
 vaidade emo:orgulho
 agressividade emo:fúria
 alegria emo:feliz_emo:satisfeito
 carinho emo:amor
 desgosto
 emo:infeliz_emo:insatisfeito
 insatisfação
 emo:infeliz_emo:insatisfeito
 prazer emo:feliz_emo:satisfeito
 ultraje emo:vergonha
 vergonha emo:vergonha

GRUPO 2

abandono	alheamento	anti-lusitanismo
abatimento	alienação	antipatia
abuso	ambiguidade /	anulação
aceitação	ambigüidade	apoio
acomodação	ambivalência	aprendizagem
adesão	aniquilamento	apropriação
adoção	anonimato	aprovação
afiliação	antagonismo	aproximação
afirmação	anticomunismo	aptidão
ajuda	antilhanía	arte

atraso	competência	desagregação
ausência	competitividade	descompromisso
austeridade	compromisso	desconforto
auto-afirmação	comunidade	desconsideração
autocontrole	concordância	descrédito
auto-depreciação	confiabilidade	desempenho
auto-identificação	conflito	desenraizamento
autorealização / auto-realização	conformação	desenvolvimento
auto-responsabilidade	confraternização	desintegração
autoritarismo	consideração	deslocamento
auto-suficiência	consolidação	desmotivação
autovalor	contestação	desordem
autovalorização / auto-valorização	continuidade	desorientação
aventura	contradição	despedida
avó	contrição	desperdício
bairrismo	controle	despersonalização
beleza	cooperação	despotismo
bem	co-propriedade	desprofissionalização
brasilidade	co-responsabilidade	desproteção
calor	corpo	desqualificação
camaradagem	corrupção	desterro
capacidade	credibilidade	destino
castração	criança	destruição
cercos	cruzada	desvalia
cidadania	culpabilidade	desvalorização
civismo	dano	dever
clã	decência	diferenciação
claustrofobia	defesa	dificuldade
coerência	degradação	discordância
coesão	déjà=vu	discriminação
colaboração	demanda	disposição
coleguismo	democracia	distância
coletividade	dependência	distanciamento
coletivo	deriva	distinção
companhia	derrota	dívida
	derrotismo	doença
	desafio	domínio

duração	falhanço	inadequação
educação	família	inanimidade
emancipação	fanatismo	incapacidade
emergência	fatalidade	inclusão
eminência	fatalismo	incomodidade
emotividade	federação	incompatibilidade
empobrecimento	festa	incompetência
enfado	filiação	incompreensão
eqüidade	fim	incongruência
equilíbrio	final	indefinição
equipe	finitude	independência
escândalo	fome	indeterminação
escassez	fracasso	individualidade
esforço	fragilidade	individualismo
especialidade	frieza	indivisibilidade
espera	fuga	inexistência
espiritualidade	fugacidade	infallibilidade
espoliação	fusão	infância
estrangeiro	futilidade	infantilização
estranheza	futuro	ingenuidade
esvaziamento	germanidade	ingovernabilidade
eternidade	germanismo	inibição
ética	governabilidade	injustiça
etnicidade	grandiosidade	insuficiência
evasão	gratificação	integração
evidência	gratuidade	integridade
exaltação	hierarquia	interdependência
excelência	identificação	intimidade
exclusão	ilegitimidade	inutilidade
exclusividade	imersão	invasão
exílio	imortalidade	inviabilidade
existência	importância	invulnerabilidade
exposição	impossibilidade	irmandade
êxtase	impotência	irrealidade
exterioridade	impunidade	irredutibilidade
faceirice	imunidade	irrelevância
falência	inadaptação	isolamento

italianidade	necessidade	povo
lealdade	nobreza	preconceito
legitimidade	novidade	premonição
liberação	obediência	presença
libertação	obrigação	preservação
liderança	obrigatoriedade	preterimento
ligação	obviedade	prevenção
localidade	omnipotência /	prisão
logro	onipotência	privação
luta	onipresença	privacidade
mãe	ônus	produtividade
maioria	oposição	progresso
mal-estar	orfandade	propósito
marginalidade	organização	propriedade
marginalização	parentesco	protesto
massa	participação	provisoriedade
maternidade	partilha	proximidade
memória	passado	pureza
mérito	passividade	qualidade
mineiridade	paternidade	quebra
minoridade	pátria	raça
mistério	patriota	racismo
mobilização	paz	razão
monotonia	pequenez	reação
moralidade	perdição	realidade
motivação	perigo	receptividade
mudança	permanência	recusa
mulher	persecutoriedade	referência
mundo	perseguição	religiosidade
nação	personalidade	renovação
nacionalidade	pertença	renúncia
nacionalização	pertencimento	representatividade
não-existência	pertinência	resistência
não-participação	pobreza	revanchismo
nativismo	poder	risco
navrágio	política	saber
náusea	posse	sacrifício

saturação	tortura	urgência
self	totalidade	utilidade
semelhança	trabalho	vácuo
separação	tradição	valor
serenidade	tragicidade	valorização
seriedade	traição	verdade
singularidade	transcendência	vertigem
soberania	transformação	vinculação
sobrecarga	transitoriedade	violência
sociabilidade	ufanismo	vitalidade
subalternidade	união	vítima
subalternização	unificação	vitimização
suspeição	universalidade	vitória

agradecimento emo:grato	gozo emo:satisfeito
agrado emo:satisfeito	heroísmo emo:coragem
altivez emo:orgulho	idolatria emo:amor
ambição emo:desejo	modéstia emo:humildade
amigo emo:amor 0	pêsames emo:infeliz
autoestima / auto-estima emo:amor	realização emo:satisfeito
bem-estar emo:feliz	rebaixamento emo:vergonha
choro emo:infeliz	reconhecimento emo:grato
decoro emo:vergonha	reverência emo:admirar
desejabilidade emo:desejo	saciedade emo:satisfeito
deslumbramento emo:admirar	ufania emo:orgulho
dignidade emo:orgulho	veneração emo:amor_emo:admirar
fé emo:esperança	vontade emo:desejo

Anexo 2

O Anexo 2 documenta na seção (a) os lemas da cauda longa que foram inseridos em grupos do léxico do Emocionário; na seção (b) os lemas que foram agrupados em campos lexicais; e na seção (c) os lemas automotivados e suas respectivas definições de anotação semântica.

(a) Lemas da cauda longa que foram inseridos em grupos do léxico do Emocionário:

AUSÊNCIA

abstenção; anestesia; inatividade; indiferença; inoperância; neutralidade; prostração; torpor; vazio

(b) Lemas agrupados em campos lexicais distintos do

PERTENCIMENTO / COLETIVIDADE

abrigo; acrianidade; aldeia; americanismo; amparo; ancestralidade; angolanidade; brasileiro; caserna; catalanidade; colectivismo; convivialidade; cooperativismo; enraizamento; escravidão; *ghetto*; goianidade; gueto; hispanidade; jacobinismo; natividade; nordestinidade; nordestino; origem; patriotismo; paulistanidade; pernambucalidade; portugalidade; *pueblo*; reapropriação; regionalidade; *regionalism*; sampaulinismo; unicidade; vizinhança

NÃO PERTENCIMENTO / ESTRANHAMENTO

afastamento; antiamericanismo; anti-germanismo; clausura; deformação; degredo; desabrigo; desapropriação; descolamento; despatriotização; despersonalização; diáspora; distanciação; estrangeiridade; estrangeirismo; impertença; não-lugar; não-reconhecimento; *outsider*; peixe “fora d’água”; prisioneiro; xenofobismo

ESPIRITUALIDADE

beatitude; católico; perdão; religião

FAMILIA

bastardia; familiaridade; filha; *paternalism*; viúva; viuvez; viúvo

(c) lemas automotivados

Os sintagmas foram interpretados caso a caso dentro do contexto no qual estavam inseridos

Três resoluções:

O lema foi anotado como emoção apenas no contexto da entrada “sentimento de auto-N”;

O lema “auto-N” foi anotado como emoção (não sendo necessário, portanto, anotar o sintagma);

O lema sem o prefixo “auto” também foi inserido em um grupo do léxico.

Lema	Sema	Lema sem o prefixo “auto”	Sema	Resolução
autosuficiência	0	suficiência	0	“sentimento de autosuficiência” = grupo Esperança “autosuficiência” e “suficiência” = mantém o sema 0
auto-repugnância	0	repugnância	0	“auto-repugnância” e “repugnância” = grupo Ódio
auto-respeito	grupos Humildad e e Admirar	respeito	grupos Humildad e e Admirar	“auto-respeito” e “respeito” = foram mantidos nos grupos Humildade e Admirar
auto-reprovação	0	reprovação	0	“auto-reprovação” e “reprovação” = grupo Vergonha
auto-reflexão	0	reflexão	0	“sentimento de auto-reflexão” = grupo Outra “reflexão” = mantém o sema 0
auto-referência	0	referência	0	“sentimento de auto-referência” = Vergonha “referência” = mantém o sema 0
auto-punição	0	punição	0	“sentimento de auto-punição” = grupo Vergonha “punição” = mantém o sema 0
autoproteção	0	proteção	0	“sentimento de autoproteção” = grupo Esperança “proteção” = mantém o sema 0
auto-preservação e autopreservação	0	preservação	0	“sentimento de autopreservação” = “preservação” = mantém o sema 0
autopiedade	Pena	pieidade	Pena	“autopiedade” e “pieidade” = mantidos no grupo Pena

Lema	Sema	Lema sem o prefixo "auto"	Sema	Resolução
autoculpabilização	0	culpabilização	0	"autoculpabilização" = grupo Vergonha "culpabilização" = mantém o sema 0
autoflagelação	0	flagelação	0	"autoflagelação" e "flagelação" = grupo Desespero
auto-indulgência	0	indulgência	0	"auto-indulgência" = grupo Alívio "indulgência" = sema 0
autoexigência	0	exigência	0	"sentimento de autoexigência" = grupo Orgulho "exigência" = mantém sema 0
auto-desvalorização	0	desvalorização	0	"auto-desvalorização" = grupo Vergonha "desvalorização" = mantém sema 0
autodepreciação	0	depreciação	0	"autodepreciação" = grupos Insatisfação e Vergonha "depreciação" = mantém sema 0
autodomínio	0	domínio	0	"sentimento de autodomínio" = grupo Alívio "domínio" = mantém sema 0
autodeterminação	0	determinação	coragem	"autodeterminação" = grupo Coragem "determinação" = mantido no grupo Coragem
autodesvalia	0	desvalia	0	"autodesvalia" = grupos Insatisfação e Vergonha "desvalia" = mantém sema 0
auto-destruição e autodestruição	0	destruição	0	"sentimento de autodestruição" = grupos Infelicidade e Vergonha "destruição" = mantém sema 0
autocrítica	0	crítica	0	"sentimento de autocrítica" e "crítica" = grupo Insatisfação
autodesmerecimento	0	desmerecimento	0	"autodesmerecimento" = grupo Vergonha "desmerecimento" = mantém sema 0
autocontraponto	0	contraponto	0	"sentimento de autocontraponto" = grupo Outra "contraponto" = mantém sema 0

Lema	Sema	Lema sem o prefixo "auto"	Sema	Resolução
autoconciliação	0	conciliação	0	"autoconciliação" e "conciliação" = grupo Alívio
autocompaixão	Pena	compaixão	pena	"autocompaixão" e "compaixão" mantidos no grupo Pena
autocobrança	0	cobrança	0	"autocobrança" = grupo Fúria "cobrança" = mantém sema 0
auto-avaliação	0	avaliação	0	"sentimento de auto-avaliação" = grupo Outra "avaliação" = mantém sema 0
auto-apreciação	0	apreciação	0	"auto-apreciação" e "apreciação" = grupos Admirar e Satisfação
auto-aplauso	0	aplauso	0	"sentimento de auto-aplauso" = grupo Orgulho "aplauso" = mantém sema 0
auto-agressividade	Fúria	agressividade	Fúria	"sentimento de auto-agressividade" e "agressividade" mantidos no grupo Fúria
autoafirmação	0	afirmação	0	"sentimento de autoafirmação" = grupo Esperança "afirmação" = mantém sema 0
autosatisfação	Felicidade e Satisfação	satisfação	Felicidade e Satisfação	"autosatisfação" e "satisfação" = mantidos nos grupos Felicidade e Satisfação
auto-complacência	0	complacência	0	"auto-complacência" = grupo Orgulho "complacência" = mantém sema 0

Tabela A - "Ato auto motivados".

Anexo 3

A Tabela lista as expressões anotadas no Esqueleto seguidas dos grupos de emoção às quais foram atribuídas. Caso a anotação do primeiro grupo já seja “0”, significa que essa expressão não foi adicionada a nenhum dos grupos de emoção. A última coluna mostra a anotação semântica de cada expressão anterior ao processo de inserção aos grupos do léxico do Emocionário.

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
à flor da pele	#genérica				sentimento
a sangue frio	#coragem				sentimento
a unhas e dentes	#coragem				sentimento
abaixar a orelha	#vergonha				sentimento
abrir o coração	#humildade				sentimento
abrir os olhos	0				sentimento
amigo do coração	#amor				sentimento
amolecer o coração	#genérica				sentimento
andar mal das pernas	0				opinião
apertar o coração	#desespero	#infelicidade			sentimento
aperto na/de garganta	#infelicidade	#desespero			sentimento
aperto no estômago	#fúria	#infelicidade			sentimento
aperto no peito	#medo	#infelicidade	#desespero		sentimento
arrancar os cabelos	#desespero				sentimento
arrepiar os cabelos	#medo				sentimento
ataque de nervos	#medo				sentdoença
bater o coração	#desejo				sentimento
bicho de sete cabeças	0				opinião
bom da cabeça	0				opinião
bom de coração	0				opinião
botar a cabeça no lugar	#alívio				sentimento
brilhar os olhos	#desejo	#admirar	#felicidade		sentimento
brilho nos olhos	#desejo	#admirar	#felicidade		sentimento
cabeça aberta	0				opinião
cabeça dura	0				opinião

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
cabeça fervendo	#fúria				sentimento
cabeça fraca	0				opinião
cabeça fresca	#alívio				opinião
cabeça fria	#alívio				sentimento
cabeça quente	#fúria				sentimento
cair o queixo	#surpresa				sentimento
cara amarrada	#insatisfação				sentimento
cara de carago	0				opinião
cara de pau	#coragem				opinião
cara de poucos amigos	#insatisfação				sentimento
cara de tacho	#vergonha				sentimento
cara fechada	#insatisfação				sentimento
caralho	#generica				sentimento
chegar aos calcanhares de	0				sentimento
chegar aos pés de	0				sentimento
chorar de barriga cheia	#insatisfação	#ingratidão			sentimento
chorar no ombro	#infelicidade				sentimento
colocar a cabeça no lugar	#alívio				sentimento
com a cara e a coragem	#coragem				sentimento
com a corda no pescoço	#desespero				sentoutros
com cara nova	0				opinião
com o cabelo em pé	#desespero				sentimento
com o coração na mão	#pena				sentimento
com o estômago na boca	#surpresa	#desespero	#medo		sentimento
com o peito aberto	#gratidão	#esperança	#humildade		sentimento
com pedras na mão	#insatisfação	#fúria			sentimento
com quatro pedras na mão	#ódio	#fúria			sentimento
com unhas e dentes	#coragem				sentimento
comer com os olhos	0				opinião
comer de joelhos	0				sentimento
coração aberto	#gratidão	#humildade			sentimento
coração ao pé da boca	#desespero	#medo			sentimento
coração apertado	#pena	#insatisfação			sentimento
coração da equipe	0				sentimento
coração de manteiga	#pena				sentopi
coração de ouro	0				sentopi

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
coração de pedra	#ausência				sentopi
coração estar	#genérica				sentimento
coração grande	0				sentopi
coração mole	#pena				sentopi
coração muito grande	0				sentimento
coração romântico	#amor				sentimento
coração saindo pela boca	#surpresa	#desespero	#medo		sentimento
cortar o coração	#pena	#infelicidade			sentimento
crise de nervos	#medo				sentdoença
custar os olhos da cara	0				opinião
dar de ombro	#ausência	#vergonha			sentimento
dar nos nervos	#fúria				sentimento
de boca aberta	#admirar	#surpresa			sentimento
de braço cruzado	#orgulho				sentimento
de braços abertos	#gratidão				sentimento
de cabeça erguida	#orgulho	#coragem	#satisfação		sentposiçã o
de cabeça levantada	#orgulho	#coragem	#satisfação		sentposiçã o
de cabelo em pé	#outra				sentimento
de cara fechada	#insatisfação				sentopi
de cara nova	0				opinião
de coração a coração	0				sentimento
de corpo e alma	#amor				sentimento
de cortar o coração	#pena				sentopi
de joelhos	#vergonha				opinião
de orelha em pé	0				sentimento
de ouvido em pé	0				sentimento
de pé	0				sentimento
de peito aberto	#gratidão	#esperança	#humildade		sentimento
de queixo caído	#surpresa				sentimento
deixar sem fôlego	#admirar				sentimento
desmiolado	0				opinião
destruidor de corações	0				opinião
do caralho	#genérica				sentimento
doente da cabeça	0				opinião
doente do pé	0				opinião
doido da cabeça	0				opinião
dor de cotovelo	0				sentimento
duro de ouvido	#orgulho				opinião
em pé de guerra	#ódio	#insatisfação	#fúria		sentimento
embrulhar o estômago	#insatisfação	#surpresa			sentopi

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
encher o saco	#insatisfação				sentopi
encher os olhos	#desejo				opinião
encolher os ombros	#vergonha				sentimento
entalar a garganta	#outro				sentimento
esfriar a cabeça	#alívio				sentimento
esquentar a cabeça	#outro				sentimento
estar de pé atrás	#desespero				sentimento
falar de alguém pelas costas	0				sentimento
falar mal pelas costas	0				sentimento
falar pelas costas	0				sentimento
falar pelos cotovelos	0				opinião
fechar a cara	#insatisfação	#infelicidade			sentimento
fechar os olhos	0				sentimento
ficar com um pé atrás	#desespero				sentimento
ficar de boca aberta	#surpresa	#admirar			sentimento
ficar sem fôlego	#surpresa				sentimento
franzir o sobrolho	#outro				sentimento
frio na barriga	#surpresa				sentimento
fundo do coração	#amor	#desejo	#gratidão	#esperança	sentimento
garganta entalada	#desespero				sentimento
gelar a espinha	#medo				sentimento
guerra de nervos	#genérica				sentimento
ir com a cara	#amor	#satisfação			sentimento
jogar na cara	#coragem	#insatisfação	#fúria		sentimento
levantar as mãos à cabeça	#surpresa				sentimento
levar as mãos à cabeça	#surpresa				sentimento
mais olhos que barriga	#desejo				sentopi
maluco da cabeça	0				opinião
manter a cabeça no lugar	#alívio				sentimento
mão aberta	0				opinião
mão fechada	0				opinião
más línguas	#inveja				opinião
meia boca/meia-boca	#insatisfação				opinião
menina dos olhos	0				opinião
morar no coração	#amor				sentimento
na cara	#coragem				sentimento
nervos	#genérica				sentfaculda de

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
nervos à flor da pele	#medo				sentimento
nervos de aço	#coragem				sentopi
nó na garganta	#desespero				sentimento
nosso coração querer	#genérica				sentimento
ódio no coração	#ódio				sentimento
olhar com bons olhos	0				opinião
olho gordo	#inveja				sentimento
olho grande	#inveja				sentimento
olho no olho	#coragem				sentimento
olhos brilhando	#desejo	#admirar	#felicidade		sentimento
olhos maiores que a barriga	#desejo				sentopi
ombro amigo	#amor				sentimento
osso duro de roer	0				opinião
partir o coração	#pena				sentimento
pé atrás	#desespero				sentimento
pé de valsa	0				opinião
pé e mão atado	#desespero				sentimento
pé frio	0				opinião
pé no saco	0				opinião
pé quente	0				opinião
pé rapado	0				opinião
perder a cabeça	#desespero				sentimento
pôr a cabeça no lugar	#alívio				sentimento
pulga atrás da orelha	#desespero				sentimento
puro de coração	0				sentopi
puxa – saco / puxa saco	#inveja				opinião
puxar o saco	0				sentopi
reclamar de barriga cheia	#insatisfação	#ingratidão			sentimento
rei na barriga	#orgulho				sentopi
rir com os olhos	#felicidade	#gratidão			sentimento
ruim da cabeça	0				opinião
saltar ao ouvido	#surpresa				sentimento
saltar aos olhos	0				sentimento
sangue ferver	#fúria				sentimento
sem fôlego	#surpresa	#admirar			sentimento
sem pé	0				sentimento
sem pé e sem cabeça	0				opinião
sem pés nem cabeça	0				opinião
sem saco	#fúria				sentimento
sentir na pele	#genérica				sentimento

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
ser coração bom	#genérica				opinião
ser de coração	#orgulho	#amor			sentimento
ser do caralho	0				opinião
ser um coração bom	#genérica				opinião
ser um tapa na cara	0				opinião
ser uma dor de cabeça	0				opinião
soco no estômago	#surpresa				opinião
sofrer dos nervos	#desespero				sentdoença
sofrer na pele	#genérica				sentimento
soltar fogo pelas ventas	#furia				sentimento
sorrir com os olhos	#felicidade	#gratidão			sentimento
tapa na cara	#surpresa				opinião
tapar os olhos	0				sentimento
ter a cabeça no lugar	0				opinião
ter bom coração	#genérica				opinião
ter bom olho	0				opinião
ter cara de	0				opinião
ter coração bom	#genérica				opinião
ter estômago	#coragem				sentimento
ter mão	0				opinião
ter o coração bom	#genérica				opinião
ter o pé atrás	#desespero				sentimento
ter papas na língua	0				opinião
ter um bom coração	#genérica				opinião
ter um coração	#genérica				opinião
ter um coração bom	#genérica				opinião
ter um coração de	#genérica				sentimento
ter um coração de manteiga	#pena				sentopi
ter um coração de ouro	0				sentopi
ter um coração de pedra	#ausência				sentopi
ter um coração duro	#orgulho				sentopi
ter um coração enorme	0				sentopi
ter um coração grande	0				sentopi
ter um coração mole	#pena				sentopi
ter um pé atrás	#desespero				sentimento
ter uma cara própria	0				opinião
time do coração	#amor				sentimento
tiro no pé	0				opinião
torcer o nariz	#insatisfação				sentopi
trazer no olho	#genérica				sentimento
trocar os pés pelas mãos	0				opinião

Expressões	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Sema
um olho aberto e outro fechado	#desespero				sentimento
unha de fome	0				sentopi
ver com bons olhos	0				opinião
ver pelas costas	0				sentimento
vergonha na cara	#vergonha				sentimento
visceral	0				opinião
viver de joelhos	#vergonha				opinião

Tabela B - Esqueleto + Emocionário.

Anexo 4

Nas seções “a” a “g” as mudanças estão documentadas para cada um dos grupos separadamente, com observações e informações extras quando necessário.

a) Grupo Desespero

Adicionados: “angustiado”, “atormentar”, “desconfiado”, “desesperançoso”, “tormento”.

Excluídos: “alvorçado”, “alvorçar”, “alvorço”, “arrebatar”, “arrebatar-se”, “desesperativo”, “insinuar”.

b) Grupo Esperança

No grupo Esperança foram feitas adições, exclusões e mudanças. Dentre as adições está o lema “crer”, que não fazia parte do Emocionário; o lema “promissoramente”, que já estava anotado como emoção, mas não listado no Emocionário; e os lemas “expetante”, “expetação” e “expectação”, que já estavam listados no Emocionário, mas sem anotação de emoção no corpus. O lema “augurar” foi excluído do léxico.

As mudanças aconteceram para os lemas “crença” e “esperar”. As regras para “crença” foram adicionadas ao VISL, já que “crença” tem o sentido emocional quando combinada com verbos no subjuntivo ou com o adjetivo inabalável:

```
ADD (%EMO:ESPERANCA) TARGET ("crença") (1 ("em")) (2 ("que"))  
(3*(SUBJ)); ADD (%EMO:ESPERANCA) TARGET ("crença") (1 ("inabalável"));
```

Para “esperar” foram feitas regras negativas focadas nos complementos do verbo, para quando “esperar” carrega o sentido de “aguardar”, não emocional. Os seguintes contextos foram excluídos da anotação de emoção:

```
a:[lema="esperar"] [lema="por(\+.*)*"] >> a:[sema="0"]  
a:[lema="esperar"] [lema="para"] >> a:[sema="0"]  
a:[lema="esperar"] [pos="NUM.*"]  
[lema="hora|minuto|segundo|ano|século|década"] >> a:[sema="0"]  
a:[lema="esperar"] [pos="DET_quant.*"] [lema="tempo"] >> a:[sema="0"]  
a:[lema="esperar"] [lema="um"] [lema="filho|filha|bebé|criança"] >>  
a:[sema="0"]
```

a:[lema="esperar"] [lema="filho|filha|bebê|bebê|criança"] >> a:[sema="0"]
[lema="esperar"] [lema="o|um"]*

@[lema="ônibus|autocarro|camioneta|comboio|trem|avião|voo|táxi"]

c) Grupo Humildade

Excluídos: “pedido”, “pedir”

Adicionado: “pedido de desculpas”

d) Grupo Pena

Adicionados: “clemência”, “clemente”, “condolência” e “compassivo”

e) Grupo Satisfação

Ao grupo Satisfação foram adicionados os lemas “apreciar”, “contentar”, “deleitar”, “deleitar-se” e “júbilo”. Além disso, havia muitos casos do lema “deleite” com sema “0” mas que indicavam emoção que foram corrigidos.

Neste grupo, algumas regras para nomes próprios e MWEs foram criadas. Foi feita uma regra para “Doutores da Alegria”, uma organização sem fins lucrativos cujos membros visitam crianças e adolescentes em hospitais fantasiados de palhaço:

a:[word="Doutores"] b:[word="da"] c:[word="Alegria"] >> <mwe sema="0"
lema="Doutores=da=Alegria" pos="N"> a: b: c: </mwe>

Partindo do lema “alegre”, foram criadas regras para nomes próprios, em sua maioria lugares no Brasil, sendo eles “Buriti Alegre”, “Córrego Alegre”, “Jardim Alegre”, “Monte Alegre”, “Porto Alegre”, “Pouso Alegre”, “Rio Alegre”, “River Alegre”, “Várzea Alegre” e “Vista Alegre”:

a:[word="Buriti"] b:[word="Alegre"] >> a:[lema="Buriti=Alegre"]

b:[lema="Buriti=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Córrego"] b:[word="Alegre"] >> a:[lema="Córrego=Alegre"]

b:[lema="Córrego=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Jardim"] b:[word="Alegre"] >> a:[lema="Jardim=Alegre"]

b:[lema="Jardim=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Monte"] b:[word="Alegre"] >> a:[lema="Monte=Alegre"]

b:[lema="Monte=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Porto"] b:[word="Alegre"] >> a:[lema="Porto=Alegre"]

b:[lema="Porto=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Pouso"] b:[word="Alegre"] >> a:[lema="Pouso=Alegre"]
b:[lema="Pouso=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Rio"] b:[word="Alegre"] >> a:[lema="Rio=Alegre"]
b:[lema="Rio=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="River"] b:[word="Alegre"] >> a:[lema="River=Alegre"]
b:[lema="River=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Várzea"] b:[word="Alegre"] >> a:[lema="Várzea=Alegre"]
b:[lema="Alto=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

a:[word="Vista"] b:[word="Alegre"] >> a:[lema="Vista=Alegre"]
b:[lema="Vista=Alegre" & pos="PROP" & sema="0"] # CBR, DHBB, NSC, CHV e FLOR

Também foi criada uma regra para “Mocidade Alegre”, uma escola de samba da cidade de São Paulo:

a:[lema="Mocidade"] b:[lema="Alegre"] >> a:[lema="Mocidade=Alegre"]
b:[lema="Mocidade=Alegre" & pos="PROP" & sema="0"] # CHV

Após as regras, há algumas marcações com *hashtag* que se referem aos corpora nos quais as regras serão válidas. A separação por corpus ocorre para marcar apenas os casos em que a combinação realmente se refere a um nome próprio. No caso de mocidade alegre, a combinação pode se referir ao período da juventude e não só à agremiação, como mostram os exemplos a seguir:

par=CdF-Mais_Uma-107: Reagiu com a sua mocidade alegre e descuidosa .

: A esse respeito, artigo anônimo intitulado «às Sextas Feiras» e publicado no Jornal do Comercio em 12 de junho de 1931, quando da ocasião da morte de Vicente Licínio, é sintomático:

Não teve, como a generalidade dos jovens, uma mocidade alegre, expansiva, despreocupada: Na Escola Politécnica, durante o curso acadêmico, alheou-se por completo do meio ambiente, em que por forças das circunstâncias deveria ter-se integrado, para dedicar-se tão apenas ao cumprimento dos seus deveres de estudante (Artigo anônimo, Jornal do Comércio, 1931: arquivo particular)

Após a leitura das linhas de concordância na distribuição por corpus, foi possível perceber que a referência à escola de samba Mocidade Alegre só estava presente no corpus CHAVE. Logo, a anotação foi criada apenas para esse corpus.

O lema “jubilação” é sinônimo de “alegria” e “contentamento intenso” como manifestação de sentimento; mas também significa “aposentadoria” ou “desligamento de um estudante de escola ou universidade” como usos que não carregam emoção. Como só havia 176 ocorrências no corpus, as linhas de concordância foram analisadas uma a uma e as seguintes regras foram criadas para marcar quando “jubilação” é utilizada pelo falante do português com sentido emocional:

```
[lema="puro|grande"] a:[lema="jubilação"] >>
a:[sema="emo:feliz_satisfeito"]
[lema="estado"] [word="de"] a:[lema="jubilação"] >>
a:[sema="emo:feliz_satisfeito"]
[lema="fator"] [word="a"] [word="mais"] [word="de"] a:[lema="jubilação"]
>> a:[sema="emo:feliz_satisfeito"]
a:[lema="jubilação"] [lema="cinéfila|dramática|intelectual|irônica|triste"] >>
a:[sema="emo:feliz_satisfeito"]
[word="a"] [word="intensidade"] [word="e"] [word="a"] a:[lema="jubilação"]
>> a:[sema="emo:feliz_satisfeito"]
[word="de"] a:[lema="jubilação"] [word="e"] [word="de"] [word="alívio"]>>
a:[sema="emo:feliz_satisfeito"]
```

f) Grupo Saudade

O lema “saudosismo” foi adicionado ao grupo Saudade, e algumas anotações que apareciam erradas para o lema “falta” foram corrigidas.

g) Grupo Surpresa

Ao grupo Surpresa foram adicionados os lemas “maravilhado” e “maravilhar” e excluídos os lemas “rompante”, “repentino” e “súbito”.