

Descrição de uma metodologia desenvolvida para revisão de um léxico de palavras de emoção

Barbara C. Ramos

PPGEL, PUC-Rio
Rua Marquês de São Vicente, 225 Rio de Janeiro, Brasil
barbaracmpramos@gmail.com

***Abstract.** This article describes a methodology that was developed to review Emocionário, a lexicon of emotion words in Portuguese from Linguateca's AC/DC project. This first portion of the review was carried out in seven of the 24 groups of Emocionário, being them “Desespero”, “Esperança”, “Humildade”, “Pena”, “Satisfação”, “Saudade” and “Surpresa”. The methodology is then illustrated, making use of examples taken from the corpus itself for each of the steps. Some of the main changes are documented and discussed at the end.*

***Resumo.** Este artigo tem por objetivo descrever a metodologia desenvolvida para revisar o Emocionário, léxico de palavras de emoção do projeto AC/DC, da Linguateca. Esta primeira parcela da revisão foi realizada em sete dos 24 grupos do Emocionário, sendo eles “Desespero”, “Esperança”, “Humildade”, “Pena”, “Satisfação”, “Saudade” e “Surpresa”. A metodologia é descrita de forma detalhada, fazendo uso de exemplos retirados do próprio corpus para ilustrar cada passo. Ao final, algumas das principais mudanças realizadas são documentadas e discutidas.*

1. Introdução

O léxico de emoções do AC/DC (Santos e Bick, 2000) foi desenvolvido no âmbito da Gramateca inicialmente por Cristina Mota e Diana Santos (2015). Sua primeira fase de desenvolvimento teve o objetivo de aumentar a informação semântica nos corpora da Linguateca. A princípio, 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 inicial para a documentação e anotação da menção de emoção que precisaria de revisão e reavaliação. Nessa primeira fase do trabalho, uma grande quantidade de informação foi agrupada e, também, diversos problemas identificados. O objetivo principal, 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) extinguiu as subdivisões da anotação das palavras de emoção em “emomin” e “emomax”, que marcavam lemas com pelo menos um sentido de emoção e lemas com a maior parte das ocorrências no corpus indicando emoção, respectivamente. Essa fase também criou quatro novos grupos, totalizando 24

grupos de emoção¹ que comportavam aproximadamente 2.800 lemas diferentes e cerca de 8.000 formas de palavras que são usadas na língua portuguesa para descrever emoção.

A partir de 2019, o projeto de emoções do AC/DC passou a se chamar Emocionário e entrou em sua terceira fase, que consiste em reavaliar, revisar e reorganizar o léxico de emoções que já existia, criando um material de anotação ainda mais refinado e preciso. A partir das regras baseadas no Emocionário, é feita a anotação nos corpora. Essas regras podem ser mais descomplicadas, indicando que a palavra é uma emoção, ou mais complexas – que acabam ocorrendo em maioria – quando a palavra indica emoção em certos contextos. O léxico do Emocionário tem como principais características (i) a possibilidade de sobreposição de grupos, podendo um mesmo lema fazer parte de mais de um grupo de emoção (como “Felicidade” e “Satisfação”), trabalhando de forma complementar segundo os sentidos das palavras; (ii) a abrangência semântica de cada grupo, que faz com que lemas como *agitado*, *exaltado* e *furioso* pertençam ao mesmo grupo, mesmo estando distantes uns dos outros considerando “níveis de fúria”; e (iii) a inclusão de palavras independentemente de sua frequência no corpus, sinalizando que carregam emoção em algum contexto. O Emocionário é, portanto, o projeto de anotação semântica das emoções, em um sentido mais amplo, e o léxico das emoções, em um sentido mais estrito, que é o resultado final observado pelo usuário da Linguateca.

Uma primeira parcela da revisão foi realizada em 2020, em sete dos 24 grupos que compõem o léxico (“Desespero”, “Esperança”, “Humildade”, “Pena”, “Satisfação”, “Saudade” e “Surpresa”). Um dos principais objetivos 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 semiautomática do campo semântico das emoções na língua.

Nas seções 2 a 5, a metodologia é descrita detalhadamente, fazendo uso de exemplos retirados do próprio corpus no processo de aplicação do método para ilustrar os passos. Em seguida, são discutidas algumas mudanças realizadas nos grupos já revisados e apresentados comentários e resultados quantitativos.

2. Método de Revisão do Emocionário

A revisão dos grupos envolve recursos da linguística com corpus como linhas de concordância, listas de frequência, distribuição por campo semântico e lemas; alterna análises de dados nas formas quantitativa e qualitativa através de leitura distante e leitura aproximada; e conta com consulta em dicionários e ferramentas de pesquisa na internet. O primeiro movimento é a análise individual dos lemas, o passo 1, que consiste em buscar se eles já estão anotados semanticamente como emoção e, se sim, como parte do grupo em questão. Na Imagem 1 está o exemplo de *desesperador* cujas entradas em

¹ Os 24 grupos da segunda fase do léxico eram “Admirar”, “Alívio”, “Amor”, “Coragem”, “Desejo”, “Desespero”, “Esperança”, “Felicidade”, “Fúria”, “Genérica”, “Gratidão”, “Humildade”, “Infelicidade”, “Ingratidão”, “Insatisfação”, “Inveja”, “Medo”, “Ódio”, “Orgulho”, “Pena”, “Satisfação”, “Saudade”, “Surpresa” e “Vergonha”.

corpus já são parte do grupo “Desespero”. Neste caso, o processo consiste apenas em conferir as linhas de concordância e confirmar se os usos do lema condizem com a anotação para mantê-lo no grupo.

Procura: [lema="desesperador"] Distribuição de sema Corpo: os corpos todos v. 7.2 837 casos.	Procura: [lema="pedido"] Distribuição de sema Corpo: os corpos todos v. 7.2 121280 casos.
<hr/> Distribuição Houve 1 valores diferentes de sema . emo:desespero 837	<hr/> Distribuição Houve 3 valores diferentes de sema . emo:humildade 106643 0 14631 emo:humildade_act-s 6

Imagem 1. Exemplo do Passo 1 com a distribuição semântica dos lemas *desesperador* e *pedido*

Para os demais casos, que são a maioria nos grupos do Emocionário já revisados, considerando a imensa variedade semântica do português, a metodologia segue um total de sete passos que podem ser agrupados em três momentos. Esses passos são apresentados em detalhe nas seções 3, 4 e 5. No primeiro momento, realizado o passo 1, seguem-se os passos 2 e 3 que observam a contextualização dos lemas não anotados e anotados semanticamente, com ou sem emoção, visitando as linhas de concordância em que aparecem. Em um segundo momento, os passos 4 e 5 confirmam ou buscam por informações adicionais sobre os sentidos dos lemas sob análise. No terceiro momento, os passos 6 e 7 concluem o processo, que acontece a partir da identificação de novos lemas que podem ser adicionados nos grupos do léxico e da definição dos lemas que serão mantidos, retirados ou mesmo transferidos ou adicionados a outro grupo do Emocionário.

3. Análise da Anotação Semântica no Contexto dos Corpora

No primeiro momento, após selecionar o lema para análise, busco quais etiquetas semânticas ele carrega (pela opção “distribuição por anotação semântica” no AC/DC), concretizando o passo 1 da metodologia, como ilustrado na seção 2 pelo lema *desesperador*. A Imagem 1 mostra também resultados referentes ao lema *pedido*, que integra o grupo “Humildade”.

No passo 2, busco pelas linhas de concordância das entradas – ainda referentes a *pedido* – que estão anotadas como emoção (sema=”emo”), como mostra a expressão de busca² na Imagem 2, para confirmar se a anotação condiz com os contextos de uso do lema. A partir desses exemplos, fica claro que o uso do lema *pedido* não tem sentido emocional em particular.

²A expressão de busca correspondente é [lema="pedido" & sema=".*emo.*"]

No passo 3, busco exemplos por linhas de concordância das entradas que estão anotadas com outros campos semânticos ou estão sem anotação semântica (`sema="0"` e `sema!="emo"`) para confirmar os usos nos recortes apresentados pelo AC/DC. 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.

Procura: `[lema="pedido" & sema="*.emo.*"]`
Pedido de uma concordância em contexto
Corpo: os corpos todos v. 7.2

106649 ocorrências.

Número de ocorrências excessivo! Tente restringir a sua procura a menos de 5000 casos.

Concordância

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

Apresenta-se uma amostra aleatória de 5000 das 106649 ocorrências encontradas.

<p>: STJ nega a consórcio o **pedido** de liminar 18/09/97 / CONSÓRCIO / Apelo a tribunal foi feito pelo Tess STJ nega a consórcio o **pedido** de liminar da Sucursal de Brasília O STJ (Superior Tribunal de Justiça) indeferiu ontem um **pedido** de liminar do consórcio Tess, que queria assegurar o direito de participar da concorrência para a exploração da telefonia celular privada no interior de São Paulo .

par=ext917023-dias-92a-2: Aproveitando o andamento rápido proporcionado pela lebre Jos Maes (2m48,57s aos 1000m, 5m40,36s aos 2000m) , Junqueira manteve-se sempre colado aos primeiros e aproveitou-se do facto para obter o mínimo olímpico que o ano passado fálhara por tão pouco (então, 8m28,10s contra os 8m28,00 **pedidos**) .

<p>: Espero, Sr. Presidente, que os Vereadores do Município de Guarulhos usem de bom senso e votem dois **pedidos** de abertura de Comissão Especial de Inquérito -- CEI .

<p>: Merece referência, ainda, as chamadas condições de ação, as quais restam identificadas na possibilidade jurídica do **pedido**, na constatação do interesse processual e na verificação da legitimidade da parte interessada .

<p>: A Câmara de Vereadores de Pirajuí vota **pedido** de cassação do prefeito José Carlos Ortega (sem partido) na próxima quinta-feira .

Imagem 2. Linhas de concordância do lema *pedido* com anotação semântica de emoção:humildade

A partir das linhas de concordância, é possível observar que não há diferenciação nos sentidos independente da anotação, que foi inicialmente feita de forma automática. No entanto, é 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) mostra que o sentido do lema *pedido* não é emocional em nenhum dos contextos. Ilustrativamente, a partir de uma busca³ 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 do Emocionário.

4. Confirmação de Sentidos dos Lemas: pesquisas em dicionário ou ferramentas de busca na internet

O segundo momento da metodologia compreende os passos 4 e 5, que se constituem a partir de pesquisas em dicionários e/ou ferramentas de pesquisa on-line. Esses passos só precisam ser aplicados em casos nos quais não seja possível confirmar o uso do sentido emocional apenas pelas linhas de concordância, ou em casos nos quais o lema carrega muitos sentidos diferentes. O passo 4 consiste 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. Ao buscar pela definição do lema *pedido* no Dicionário Aulete Digital, também não pude confirmar o uso emocional do lema *pedido*.

³ A expressão de busca correspondente é `[lema="pedido" & sema="*.emo.*"] [lema="de"] @[pos="N"]`

Ou seja, nem dentro dos contextos das linhas de concordância, nem isolado desses contextos que o corpus nos mostra, o lema carrega o sentido emocional. 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. Ele só acontece quando o dicionário não apresenta nenhuma definição emocional isolada, mas o contexto das linhas de concordância claramente indica emoção. Como no caso do lema *pedido* não houve menção de emoção, o passo 5 não teve utilidade. Vale ressaltar que o passo 5 só é aplicado na metodologia como um último recurso ao qual, na maior parte das vezes, não precisei recorrer. Quando utilizado na revisão de outros grupos, busco na ferramenta on-line pelas frases “X é sentimento” e/ou “X é emoção”.

5. Resolução da Análise: manutenção, exclusão, adição ou transferência de lemas

O terceiro e último momento da metodologia une os passos 6 e 7. O passo 6 acontece ao longo da leitura das linhas de concordância nos passos 2 e 3. Nesse processo, observo coocorrência 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 *clémencia*, *clemente*, *compassivo* e *condolência*. Após a observação das linhas de concordância na pesquisa lema a lema no processo de revisão do grupo, essas palavras coocorriam com os lemas em análise em muitos dos exemplos observados.

Já ao observar o lema *modéstia*, parte do grupo “Humildade”, uma expressão se destacou do corpus ao longo da observação das linhas de concordância: “modéstia à parte”. Essa combinação modifica o sentido de uso do lema. Então, será criada uma regra de anotação que indique que essa Expressão de Várias Palavras (EVP) com o lema *modéstia* sai do grupo “Humildade” para o grupo “Orgulho”. Finalmente, no passo 7, mantenho ou retiro o lema previamente listado no grupo. No caso do grupo “Humildade”, após a aplicação da metodologia foram retirados os lemas *pedido* e *pedir*, que serão discutidos na seção 6.

6. Discussão

Após a aplicação da metodologia, como explicitado nos passos 1 a 4, o lema *pedido* foi retirado do Emocionário por não carregar sentido emocional. O lema *pedir* também foi excluído pelo mesmo motivo. Além do sentido emocional não estar presente nos usos, ao estarem anotados como emoção, os lemas acabam por confundir resultados de buscas por anotação semântica de emoção em pesquisas que podem se interessar justamente por menção de emoção no corpus. Por exemplo, digamos que um pesquisador queira encontrar menção de emoção no texto de Machado de Assis. A Imagem 3 mostra que as menções do grupo “Humildade” aparecem entre as dez emoções mais recorrentes em sua obra.

Procura: [autor="MacAss" & sema=".*emo.*"]
Distribuição de sema
Corpo: OBras v. 10.18

56007 casos.

Distribuição

Houve 712 valores diferentes de sema.

emo:desejo	4371	emo:humildade	1486
emo:amor	4333	emo:medo	1135
emo:feliz	3294	emo:surpresa	990
emo:infeliz	2365	emo:feliz_emo:satisfeito	845
emo:amor_Hfam	2280	emo:vergonha	736
emo:gen	1734	emo:humildade_dizer	687
emo:amor_ac_H_am	1516		

Imagem 3. Distribuição semântica de emoção na obra de Machado de Assis

Procura: [autor="MacAss" & sema=".*emo.*" & lema!="pedido|pedir"]
Distribuição de sema
Corpo: OBras v. 10.21

54486 casos.

Distribuição

Houve 706 valores diferentes de sema.

emo:desejo	4371	emo:medo	1135
emo:amor	4333	emo:surpresa	990
emo:feliz	3294	emo:feliz_emo:satisfeito	845
emo:infeliz	2365	emo:vergonha	736
emo:amor_Hfam	2280	emo:feliz_am	658
emo:gen	1734	emo:alivio	653
emo:amor_ac_H_am	1516	emo:humildade	641

Imagem 4. Distribuição semântica de emoção na obra de Machado de Assis sem os lemas *pedido* e *pedir* anotados como emoção

Na Imagem 4 aparece o resultado das menções de emoção na obra de Machado de Assis excluindo-se os lemas *pedido* e *pedir*. A ocorrência de “Humildade” cai em quase setenta por cento, retirando o grupo das dez primeiras emoções, dando mais visibilidade a outros grupos de emoção, alterando a quantidade de menção de emoção na obra do autor e, por consequência, modificando bastante os resultados da pesquisa.

A carga semântica emocional de algumas palavras é mais forte que a de outras. O verbo *pedir*, por exemplo, passou a não integrar o Emocionário porque não faz necessariamente menção a uma emoção, ao passo que os verbos *suplicar* ou *implorar* significam “pedir com súplica, humildade” ou “pedir em uma situação de desespero”. Ou seja, *suplicar* e *implorar* funcionam como a ação de *pedir* carregada com uma intenção emocional, como mostram os exemplos 01 e 02.

Exemplo 01: *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 02: *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 .

O verbo *pedir* também pode ter sua força semântica aumentada ou diminuída 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 Imagem 5 mostra o que os falantes de português pedem no corpus OBRas (Santos et al., 2018):

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	informaçõe	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	satisfaçõe	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

Imagem 5. Recorte dos complementos do lema *pedir*

A Imagem 5 faz um recorte dos cinquenta primeiros lemas de uma distribuição total de duzentos que complementam o verbo *pedir* no corpus OBRas, composto por obras literárias brasileiras. Por esse recorte, já é possível perceber a carga emocional diferente entre os complementos de *pedir*: “pedir café” não faz menção à emoção já pela leitura distante do corpus. “Pedir misericórdia” faz menção à emoção pelo uso do lema *misericórdia*. Existe, portanto, um gradiente entre usos que exprimem claramente emoções e outros que não a exprimem.

7. Comentários Finais

A flexibilidade do projeto Emocionário se manifesta exatamente neste ponto de análise da semântica dos lemas do corpus. A observação e resolução não ocorrem de forma superficial, simplesmente por determinarem que cada lema esteja ou não no Emocionário, visto que a língua também não funciona dessa forma. Em alguns casos, podem ser feitas regras para a anotação específica no corpus. O lema *alegre*, que compõe os grupos “Felicidade” e “Satisfação”, quando combinado com os lemas *Alto*, *Buriti*, *Córrego*, *Jardim*, *Monte*, *Porto*, *Pouso*, *Rio*, *River*, *Várzea* e *Vista* se refere a nomes de lugares. Logo, a anotação foi adaptada para que os casos em que *alegre* não se

refira a nomes de lugares não façam mais parte da anotação semântica de emoção em vez de ser excluído dos grupos do léxico.

Nos casos de verbos como *pedir*, uma possibilidade é analisar seus complementos tanto pelas linhas de concordância como pelas distribuições gramatical ou de lema, por exemplo, para identificar combinações recorrentes que façam menção de emoção e anotar especificamente esses casos, como no caso de “a pedido de”, expressão na qual *pedido* não tem sentido emocional, e por isso não aparece anotado como parte do grupo “Humildade”. O mesmo acontece no exemplo da expressão “modéstia à parte” na seção 5.

Tabela 1. Panorama quantitativo da revisão do léxico do Emocionário

Grupo	Lemas antes da revisão	Lemas excluídos	Lemas adicionados	Lemas adicionados a outros grupos	Lemas após a revisão
Desespero	40	4: arrebatat, arrebatat-se, desesperativo, insinuar.	8: angústia, angustiado, atormentado, atormentar desconfiado, desesperançoso, súplica, tormento.	11: amargura, atormentado, atormentar, tormento desconfiado, desconfiança, desconfiar mortificar súplica, suplicante, suplicar.	44
Esperança	36	1: crença.	2: crer, promissoramente.	2: entrever, augurar.	37
Humildade	24	2: pedido, pedir.	0	0	22
Pena	12	0	4: clemência, clemente, condolência, compassivo	0	16
Satisfação	37	1: agradado	6: apreciar, contentar, deleitar, deleitar-se, deleite, júbilo.	1: apreciar	42
Saudade	10	0	1: saudosismo.	0	11
Surpresa	41	1: súbito.	2: maravilhado, maravilhar.	0	42

Em termos quantitativos, a Tabela 1 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”. Como resultado final, os números não parecem tão distintos do cenário de antes da revisão. No entanto, por meio dessa revisão parcial do léxico foi possível mensurar a dificuldade de analisar aspectos semânticos dentro da língua com foco na descrição de emoção. A revisão também me permitiu delinear características e desafios da análise de emoção em português, 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.

Referências

- Mota, C. e Santos, D. (2015) Emotions in natural language: a broad-coverage perspective. In: Linguateca, <https://www.linguateca.pt/acesso/EmotionsBC.pdf>, Maio.
- Santos, D, Freitas, C. e Bick, E. (2018) "OBras: a fully annotated and partially human-revised corpus of Brazilian literary works in the public domain", OpenCor, Canela, RGS, Brasil, Setembro.
- Santos, D. e Bick, E. (2000) "Providing Internet access to Portuguese corpora: the AC/DC project", *Proceedings of the Second International Conference on Language Resources and Evaluation*, Editado por Maria Gavrilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis e Gregory Stainhauer, LREC 2000, págs 205–210.
- Santos, D., Simoes, A., e Mota, C. (2021) Broad coverage emotion annotation. In: *Language Resources and Evaluation*, No prelo.