

Estudo de sentimentos: algumas direções

Diana Santos* e Alberto Simões† e Cristina Mota‡
Linguatca; *Universidade de Oslo; †2Ai, EST/IPCA; ‡INESC-ID

Um domínio da experiência humana

De acordo com Wierzbicka [6], diferentes línguas conceptualizam sentimentos de forma distinta. Contudo, muitos trabalhos assumem que os sentimentos são "os mesmos" em todas as línguas, e que basta portanto traduzi-los do inglês. O nosso, não.

Sentimentos na língua

Poderíamos estudar a questão dos sentimentos de três formas distintas, como apontado por Paltoglou et al. [1]:

1. referência através da língua
2. expressão dos sentimentos em si
3. como inspirar sentimentos através da língua

O segundo campo é o objetivo do que em inglês se chama *sentiment analysis*, ou garimpo de opiniões, sobre produtos, acontecimentos ou pessoas.

O terceiro tem a ver com estudos retóricos, e inclui-se nos estudos de receção.

É o primeiro ramo que nos interessa aqui, nomeadamente a descrição de sentimentos numa dada língua: que meios lexicais, morfológicos e sintáticos permitem fazer referência ao campo semântico da emoção em português.

Como delimitar o campo semântico da emoção?

Por comparação com os campos da cor ou do corpo, o campo da emoção suscita muito mais dúvidas, porque esta não é algo objetivo (visual ou social) que pode ser confirmado por outros meios.

Mas acreditamos que é possível comunicar em português sobre emoções, o que significa que tem de haver uma intersubjetividade partilhada sobre o assunto, que permite que falantes da língua concordem sobre palavras e contextos, mesmo que haja mais discordância por se tratar de assuntos complexos e conceitos abstratos.

Tentámos portanto marcar todas as palavras e expressões que convencionalmente se referem a sentimentos em grandes quantidades de texto. Usando a nossa competência linguística (da equipa), recorrendo a dicionários e enciclopédias, e a procuras em corpos.

Essa marcação é incorporada nos corpos da Linguatca, num processo iterativo com intervenção humana, como descrito em [4], e que se encontra em curso.

Exemplos em que foram criadas regras de desambiguação:

- À palavra *admirar* podem ser atribuídos três significados diferentes (não necessariamente mutuamente exclusivos), exemplificados por: *Fiquei admirada pela sua resposta; Ela era admirada por todos; Ela admirava a estátua sempre que por ali passava.*[5]
- A palavra *reconhecer* tem vários significados, alguns referentes a sentimentos como GRATIDÃO ou VERGONHA, outros apenas associados a memória.
- A palavra *pena* pode ser associada a TRISTEZA ou a COMPAIXÃO, além de a muitos outros sentidos não emocionais.
- A palavra *sentir* é usada para sentimentos e para sensações físicas e outras.
- O mesmo se passa com a palavra *dor*, que é usada muito frequentemente de forma abstrata, às vezes vaga entre as duas interpretações

Resultados

Anotámos os corpos da Linguatca (AC/DC), que podem ser descritos como:

- Os primeiros corpos de português acessíveis para consulta através da internet (desde 1999 [3])
- Dos maiores corpos anotados do português acessíveis: mais de 1.500 milhões de palavras
- Grande variedade de géneros e de tipo de anotação [2]

Em relação à anotação, temos os seguintes dados:

Tamanho dos grupos de emoção, em palavras, e em número diferente de palavras.

alívio	admirar	amor	ausência	coragem	desejo	desespero	esperança	feliz
129365	244261	1445278	37325	360316	1971346	145101	772436	733800
262	224	971	26	442	412	157	281	828
fúria	gen	grato	humildade	infeliz	ingrato	insatisfeito	inveja	medo
753785	515838	288328	822910	707162	6102	153599	25475	614988
459	262	210	221	628	22	73	44	555
ódio	orgulho	outra	pena	satisfeito	saudade	surpresa	vergonha	
134666	370532	224329	130044	178597	106152	267579	414212	
199	565	174	216	268	80	208	657	

Grau de emocionalidade lexical

É óbvio que não conseguiremos obter uma anotação perfeita de tão grandes quantidades de texto, e ainda menos de textos que não façam parte dos corpos da Linguatca.

Por isso, é interessante arranjar formas de estimar a quantidade de referências a emoção de um dado texto sem ter de rever todas as suas ocorrências.

Para isso definimos, com base em textos já anotados e revistos, uma medida associada a cada item lexical que indica a probabilidade de ser usado no seu sentido emocional, que chamamos **grau de emocionalidade lexical**.

Alguns exemplos em texto literário: *pena* 13%, *fã* 100%, *dor* 51-60%...

Alguns exemplos em texto jornalístico: *raiva* 73%, *reconhecimento* 32%, *aterrar* 3%...

Exemplo: paleta de emoções de diferentes escritores

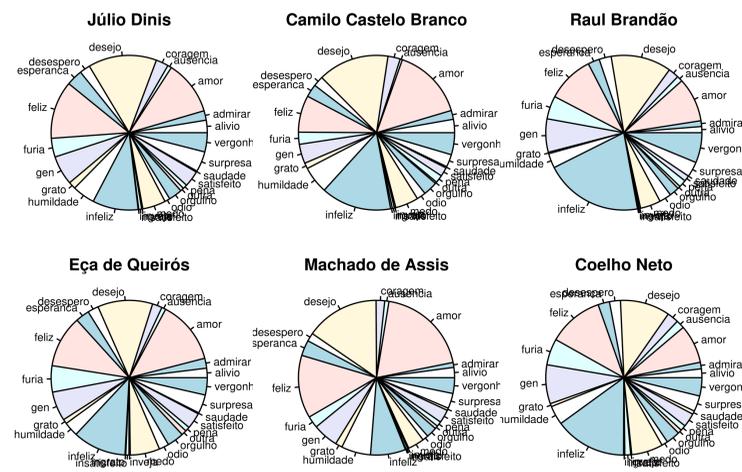


Fig. 1: Quais os escritores com maior densidade e variedade emocional?

Exemplo: rir e chorar de emoção

Um caso interessante: quais as emoções associadas ao choro e ao riso?

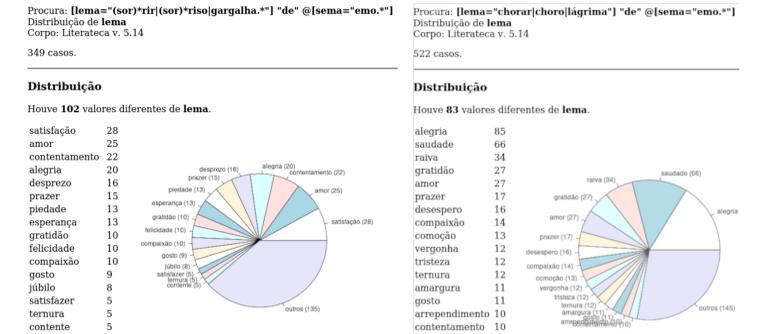


Fig. 2: Rir e chorar de emoção na literatura em português

Algumas direções

- Considerando o espaço vectorial criado pelos vários grupos de emoção, diferentes obras, diferentes géneros, diferentes autores.
- Considerando o espaço definido por palavras pulverizadas (*word embeddings*) obtidas em grandes quantidades de texto, quais as emoções mais próximas?
- O estudo das emoções ao longo do tempo, usando os textos literários

Agradecimentos

A toda a equipa da Linguatca, e sobretudo àquelas que fazem parte da equipa de anotação das emoções: Cláudia Freitas e Bárbara Ramos.

Referências

- [1] Paltoglou, Georgios, Mike Thelwall & Kevan Buckley. 2010. Online textual communications annotated with grades of emotion strength. Em *Proceedings of the 3rd international workshop of Emotion: Corpora for research on emotion and affect*, 25-31.
- [2] Santos, Diana. 2014. Corpora at Linguatca: Vision and roads taken. Em Tony Berber Sardinha & Telma Ferreira (eds.), *Working with Portuguese Corpora*, 219-236. Bloomsbury.
- [3] Santos, Diana & Eckhard Bick. 2000. Providing Internet access to Portuguese corpora: the AC/DC project. Em Maria Gavrilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis & Gregory Stainhauer (eds.), *Proceedings of LREC 2000*, 205-210.
- [4] Santos, Diana & Cristina Mota. 2010. Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora. Em N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner & D. Tapias (eds.), *Proceedings of LREC 2010*, 1437-1444.
- [5] Santos, Diana & Cristina Mota. 2015. A admiração à luz dos corpos. Em Alberto Simões, Anabela Barreiro, Diana Santos, Rui Sousa-Silva & Stella E. O. Tagnin (eds.), *Linguística, Informática e Tradução: Mundos que se Cruzam. Homenagem a Belinda Maia*, vol. 7 (1), 57-77. OSLA.
- [6] Wierzbicka, Anna. 1999. *Emotions across languages and cultures: Diversity and universals*. Cambridge University Press.