

# Estudo de sentimentos: algumas direcões

Diana Santos e Alberto Simões e Cristina Mota<sup>z</sup>  
Linguatca; Universidade de Oslo; <sup>y</sup>2Ai, EST/IPCA; <sup>z</sup>INESC-ID

## Um dom nio da experiēncia humana

De acordo com Wierzbicka [6], diferentes lnguas conceptualizam sentimentos de forma distinta. Contudo, muitos trabalhos assumem que os sentimentos so "os mesmos" em todas as lnguas, e que basta portanto traduzi-los do ingls. O nosso, no.

## Sentimentos na lngua

Poder amos estudar a questao dos sentimentos de trs formas distintas, como apontado por Paltoglou et al. [1]:

1. referēncia atraves da lngua
2. expressao dos sentimentos em si
3. como inspirar sentimentos atraves da lngua

O segundo campo e o objetivo do que em ingls se chama *sentiment analysis*, ou garimpo de opinies, sobre produtos, acontecimentos ou pessoas.

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

E o primeiro ramo que nos interessa aqui, nomeadamente a descricao de sentimentos numa dada lngua: que meios lexicais, morfolgicos e sintaticos permitem fazer referēncia ao campo semântico da emocao em portugus.

## Como delimitar o campo semântico da emocao?

Por comparacao com os campos da cor ou do corpo, o campo da emocao suscita muito mais duvidas, porque esta no e algo objetivo (visual ou social) que pode ser con rmado por outros meios.

Mas acreditamos que e poss vel comunicar em portugus sobre emocões, o que signi ca que tem de haver uma intersubjetividade partilhada sobre o assunto, que permite que falantes da lngua concordem sobre palavras e contextos, mesmo que haja mais discordância por se tratar de assuntos complexos e conceitos abstratos.

Tentamos portanto marcar todas as palavras e expressões que convencionalmente se referem a sentimentos em grandes quantidades de texto. Usando a nossa competēncia lingu stica (da equipa), recorrendo a dicionarios e enciclopedias, e a procuras em corpos.

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

Exemplos em que foram criadas regras de desambiguacao:

- A palavra *admirar* podem ser atribu dos trs signi cados diferentes (nao necessariamente mutuamente exclusivos), exempli cados por: *Fiquei admirada pela sua resposta; Ela era admirada por todos; Ela admirava a estatua sempre que por ali passava.*[5]
- A palavra *reconhecer* tem varios signi cados, alguns referentes a sentimentos como *gratidao* ou *vergonha*, outros apenas associados a memoria.
- A palavra *pena* pode ser associada a *tristeza* ou a *compaixao*, alem de a muitos outros sentidos nao emocionais.
- A palavra *sentir* e usada para sentimentos e para sensações físicas e outras.
- O mesmo se passa com a palavra *dor*, que e usada muito frequentemente de forma abstrata, as vezes vaga entre as duas interpretações

## Resultados

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

- Os primeiros corpos de portugus acess veis para consulta atraves da internet (desde 1999 [3])
- Dos maiores corpos anotados do portugus acess veis: mais de 1.500 milhees de palavras
- Grande variedade de generos e de tipo de anotacao [2]

Em relacao a anotacao, temos os seguintes dados:

### Tamanho dos grupos de emocao, em palavras, e em numero diferente de palavras.

alivio	admirar	amor	ausencia	coragem	desejo	desespero	esperanca	feliz
129365	244261	1445278	37325	360316	1971346	145101	772436	733800
262	224	971	26	442	412	157	281	828
furia	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
odio	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

E obvio que nao conseguiremos obter uma anotacao perfeita de tao grandes quantidades de texto, e ainda menos de textos que nao facam parte dos corpos da Linguatca.

Por isso, e interessante arranjar formas de estimar a quantidade de referēncias a emocao de um dado texto sem ter de rever todas as suas ocorrēncias.

Para isso de nimos, com base em textos ja 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 literario: *pena* 13%, *fa* 100%, *dor* 51-60%...

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

## Exemplo: paleta de emocões de diferentes escritores

## Exemplo: rir e chorar de emocao

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

Fig. 2: Rir e chorar de emocao na literatura em portugus

## Algumas direcões

- Considerando o espaco vectorial criado pelos varios grupos de emocao, diferentes obras, diferentes generos, diferentes autores.
- Considerando o espaco de nido por palavras pulverizadas (*word embeddings*) obtidas em grandes quantidades de texto, quais as emocões mais proximas?
- O estudo das emocões ao longo do tempo, usando os textos literarios

## Agradecimentos

A toda a equipa da Linguatca, e sobretudo aquelas que fazem parte da equipa de anotacao das emocões: Claudia Freitas e Barbara 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 admiracao a luz dos corpos. Em Alberto Simees, Anabela Barreiro, Diana Santos, Rui Sousa-Silva & Stella E. O. Tagnin (eds.), *Lingu stica, Informatica e Traducão: 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.