

Looking at Body Expressions to Enrich Emotion Clusters

Barbara Ramos^{1,2}[0000-0002-0583-5109], Diana Santos^{3,2}[0000-0002-3108-7706],
and Cláudia Freitas^{1,2}[0000-0001-6807-8558]

¹ PUC-Rio

² Linguateca

³ University of Oslo

barbaracmpramos@gmail.com, d.s.m.santos@ilos.uio.no, maclaudia@gmail.com

Abstract. This paper combines two of Linguateca’s projects – Esqueleto (on the human body) and Emocionário (on emotions), to investigate which areas of the human body and, therefore, body words or expressions, are mentioned when Portuguese speakers want to describe emotions or opinions. We present a body map as a novel research tool to describe the connections between the two domains.

Keywords: Emotions · Body · Corpus annotation · Visualization.

1 Introduction

This paper describes an attempt to cross-fertilize two projects which have used the same empirical basis (Linguateca’s AC/DC cluster) to get at corpus-based descriptions of respectively the human body and the emotional realm in Portuguese, but have so far been developed independently. Given the widely acknowledged close connections between body and emotion, it seems a natural next step to probe how these projects can illuminate each other. In this paper, we look at the following aspects:

1. from the point of view of the work on emotions, we want to investigate a) whether the already identified bodily expressions are specific to particular emotions/emotion clusters, and b) whether they are either just an expressive extension of an already identified emotion cluster or constitute rather new cases that have not yet been considered in the emotion framework, as was often the case in a previous experiment for increasing emotion coverage[3], where 90% of the lemmas found matching the pattern ”sentimento de N” (*feeling of N*) were classified as emotion.
2. from the point of view of studying the body, we want to investigate whether there is correlation between some emotions and body parts so as to understand better how the body is conceived in Portuguese, see also [8].

The research was conducted using the Gramateca framework [4, 7] and the underlying annotated corpora [6, 5].

2 Body-related emotions

In the Esqueleto project we compiled a sizeable lexicon of words and expressions containing reference to the human body, classifying them further as to whether the body was taken literally, or used with ten different meaning extensions [1].

Two of the extended meanings connected with the human body were precisely feelings and opinions, as in the case of the following examples: *estar de pé atrás* (lit. to have one foot back, mistrust), *dor de cotovelo* (lit. elbow pain, envy), *ficar com a pulga atrás da orelha* (lit. to have a flea on the back of the ear, being suspicious) for feelings (**corpo:sentimento**) or *ver com bons olhos* (lit. see with good eyes, to approve), *cabeça dura* (headstrong) for opinion (**corpo:opinia**o).

In addition there was a group of cases where in Esqueleto we had assigned both feeling and opinion, such as *encher o saco* (lit. fill the stomach, bother/be bothered), *coração de manteiga* (lit. butter heart, to have a soft heart).

Independently we also started the annotation of emotions, see [2], considering 20 and then 24 emotion groups into which to classify the ca 500 words with emotional content. This project is still being actively pursued, giving rise to a more extensive documentation of the choices in what we call *Emocionário*.

It was therefore natural to look at the body expressions already classified as emotions to see whether they would bring a more complete picture of emotion in Portuguese.

As might be expected, it was not easy to assign them to particular clusters: one thing is to agree that some emotion is involved in a particular expression, another is to identify which.

Our initial attempts can be described this way:

- We could place ca. 70% of the cases marked as emotions in one or several clusters: we have now to either create further clusters to handle the remaining emotions, or remove the emotion label
- Only a fourth of those marked as opinion could find an appropriate emotion: this we believe reinforces the need to separate between the two cases – an opinion can be uttered without emotional involvement
- The most common emotions expressed by the body expressions were dissatisfaction (**insatisfação**) and surprise (**surpresa**)

We still need to check how complex the task is and how subjective the emotion assignment task is, not least for fixed or semifixed expressions, which seem to have a wide range of meanings and interpretations, displaying even regional or variety differences in Portuguese (for example, both *dor de cotovelo* and *com o coração nas mãos* have slightly different meanings in Brazil and in Portugal. Since the first author will be developing a set of criteria to judge emotionality and emotion group membership in her Phd thesis, we will postpone exact numbers until these criteria can be applied.

3 Co-occurrence of emotions and body in context

Another study about the interrelationship of both semantic domains can be done by looking at the (more or less) free co-occurrence of words or expressions of the two domains.

Using Gramateca, a dedicated environment for doing grammar-based studies, we counted the cases where both an emotion and a body part were mentioned in the same sentence, and confirmed that these two semantic domains strongly co-occur, especially in the literary genre. In the Literateca corpus (version 3.2, ca 40 million tokens), 49.0% of the sentences with a body word also include an emotion word, while 12.9% of sentences with emotions have a body word in the same sentence. In Table 1 we present the numbers for some specific emotions:

Table 1. How often do some emotions occur, and occur with a body word, and the most common emotion word in the co-occurrences.

Emotion group	Total	With body	Relative	Most common word
gen (generic)	31,733	8,156	0.257	sentir (feel)
furia (anger)	31,091	6,914	0.222	violento (violent)
surpresa (surprise)	22,759	4,680	0.214	admirar (astonish)
vergonha (shame)	29,949	6,224	0.207	vergonha
feliz (joy)	88,662	16,908	0.191	sorriso (smile)
medo (fear)	46,567	8,810	0.189	medo
admirar (admiration)	13,900	2,776	0.164	respeito (respect)
grato (gratitude)	8,668	1,382	0.159	agradecer (thank)
ingrato (ungratefulness)	2,238	329	0.147	ingrato

Briefly, one can see that one in five expressions of a generic emotion co-occurs with a body word in the same sentence, but it is also important to note that the emotion classification has not yet been revised. So *sentir* (feel) can be about a physical sensation as well.

The other way around, we counted how often emotions are associated with one body part and created a "body map" with all related emotions, shown in Figure 1. Body parts in the map only include literal instances of body parts, classified in larger groups such as Head, Arms, Legs, Body, Internal and Hair. In Table 2 we show the numbers of sentences including a body word classified as belonging to that body part, together with the most frequent co-occurring emotion.

From a cursory analysis of 200 random instances, we could identify some recurring cases of co-occurrence:

- cases where an emotion is assigned to a part of the body, like in *olhos tristes* or *rosto triste* (sad eyes or face) or *riram-se-lhe olhos e dentes* (both his eyes and teeth laughed)
- expressions where body is related to health, and health is a positive quality along moral ones, like in *tão feliz, tão sadio de olhos*, (so happy, so healthy in his eyes).

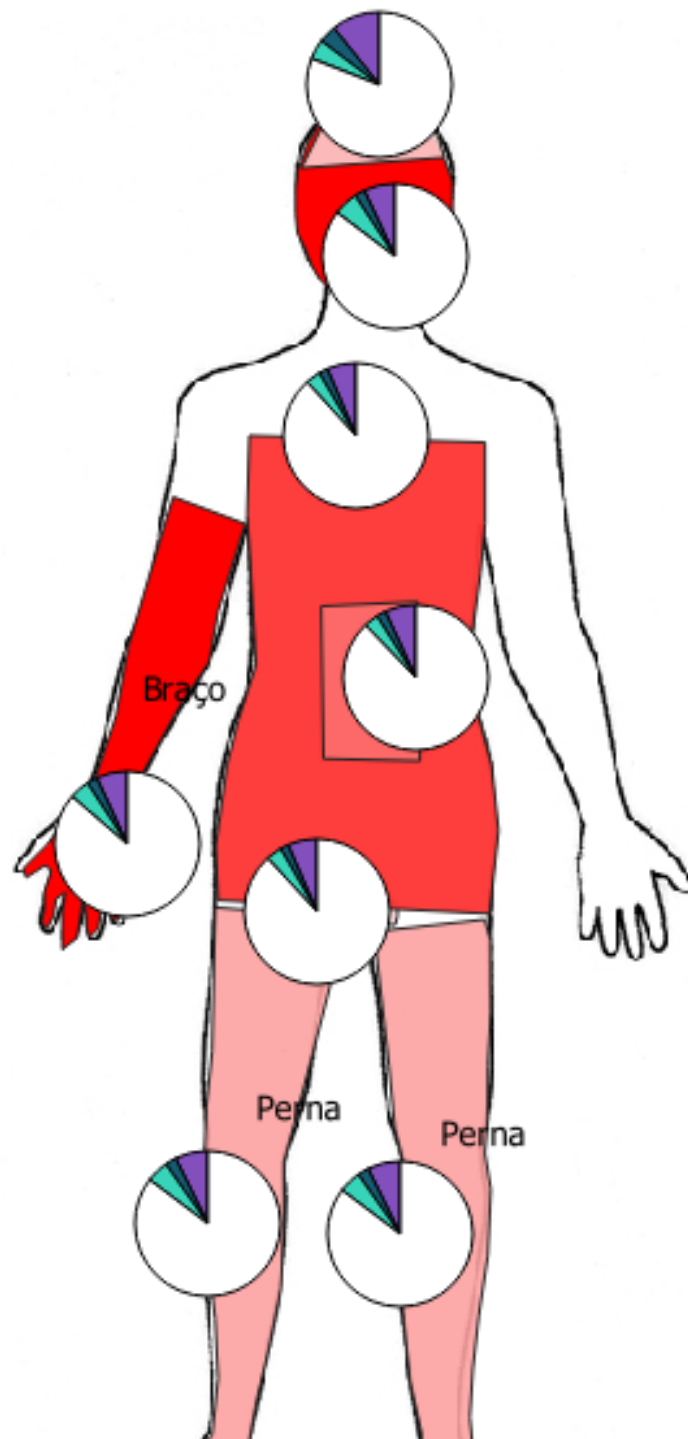


Fig. 1. The body is coloured depending on how much emotion co-occurs with it. The pie charts show, for each body part, how much comes from FURIA (rage), SURPRESA (surprise) and ADMIRAR (awe) in purple, blue and turquoise respectively.

Table 2. Body parts with emotions

Body part	size	with emotion	relative	most frequent emotion group
Cabeça (head)	82,932	40,221	0.48	feliz
Braço (arm)	48,473	21,220	0.44	amor
Tronco (body)	39,977	20,308	0.51	amor
Interno (internals)	33,222	17,324	0.52	amor
Perna (leg)	19,864	8,119	0.41	amor
Cabelo (hair)	10,327	4,420	0.43	feliz

- expressions where a body part is used metaphorically, either in standardised expressions like in *sangue derramado* (spilt blood), *entregar-lhe em mão* (give in his hand), or in specific literary descriptions such as *atirava as tranças negras (...) à cara da sociedade* (she threw her black tresses to the face of the society) or *estas lágrimas de velho mas enxuguem eles com a sua alegria* (these old man’s tears should they dry with their joy). (These should not have been classified as literally body, but not all instances of human body have been revised, which means that the results in this mapping are not precise, containing as well a large amount of free metaphorical cases.)

4 Concluding remarks

Although this is still preliminary work and we cannot present a rigorous quantitative summing up of the several experiments and explorations, we believe we are in the right path to improve both emotion annotation and evaluate it, this way getting and sharing further insights on the semantics of the Portuguese language.

We also expect that by using a methodology inspired in geographical information systems (GIS) for visualization of the interaction of body and emotion we may be able to identify further alleys for investigating the connection between these two fields, if we agree with White [9] that spatial visualization is more than just communicating results – it is a way of doing research.

References

1. Freitas, C., Santos, D., Mota, C., Carriço, B., Jansen, H.: O léxico do corpo e anotação de sentidos em grandes corpora: o projeto Esqueleto. *Revista de Estudos da Linguagem* **23**(3), 641–680 (2015)
2. Mota, C., Santos, D.: Emotions in natural language: a broad-coverage perspective. Tech. rep., Linguatca (January 2015)
3. Ramos, B., Freitas, C.: "Sentimento de quê?" uma lista de sentimentos para a Análise de Sentimentos. In: *STIL - Symposium in Information and Human Language Technology* (October, 15 - 18, 2019, Salvador, BA) (2019)
4. Santos, D.: Corpora at Linguatca: Vision and roads taken. In: Berber Sardinha, T., Ferreira, T.d.L.S.B. (eds.) *Working with Portuguese Corpora*. pp. 219–236. Bloomsbury

5. Santos, D.: Gramateca: corpus-based grammar of Portuguese. In: Baptista, J., Mamede, N., Candeias, S., Paraboni, I., Pardo, T.A., das Graças Volpe Nunes, M. (eds.) International Conference on Computational Processing of Portuguese (PROPOR'2014). pp. 214–219. Springer (October 2014), <http://www.linguateca.pt/Diana/download/gramateca.pdf>
6. Santos, D., Bick, E.: Providing Internet access to Portuguese corpora: the AC/DC project. In: Gavrilidou, M., Carayannis, G., Markantonatou, S., Piperidis, S., Stainhauer, G. (eds.) Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000). pp. 205–210 (31 May-2 June 2000)
7. Santos, D., Marques, R., Freitas, C., Mota, C., Simões, A.: Comparando anotações linguísticas na Gramateca: filosofia, ferramentas e exemplos. *Domínios de Linguagem* **10**, 11–26 (2015)
8. Sharifian, F., Dirven, R., Yu, N., Niemeier, S.: Culture and language: Looking for the "mind" inside the body, pp. 3–23. Mouton de Gruyter
9. White, R.: What is spatial history? Tech. rep., <https://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>