

Emotion-annotated corpora

can they help us detect cultural norms?

Diana Santos

d.s.m.santos@ilos.uio.no

17 February 2016



Emotion-annotated corpora

By this I mean corpora (which have been compiled for all sorts of other reasons) tagged with

- information that the lexical item mentions/refers/describes an emotion by means of its dictionary meaning
- which emotion(s) based on the assignment of all lexical items to some fields/emotion groups

Both are linguistic tasks that require interpretation, and that can only imprecisely be done automatically. Therefore, one of the points of doing this is precisely to investigate how useful this can be for the study of emotions in language: an **empirical** question.

Very briefly, what corpora

Ever since launching Linguateca we have been working to make corpora available to the public, both Web access and full download.

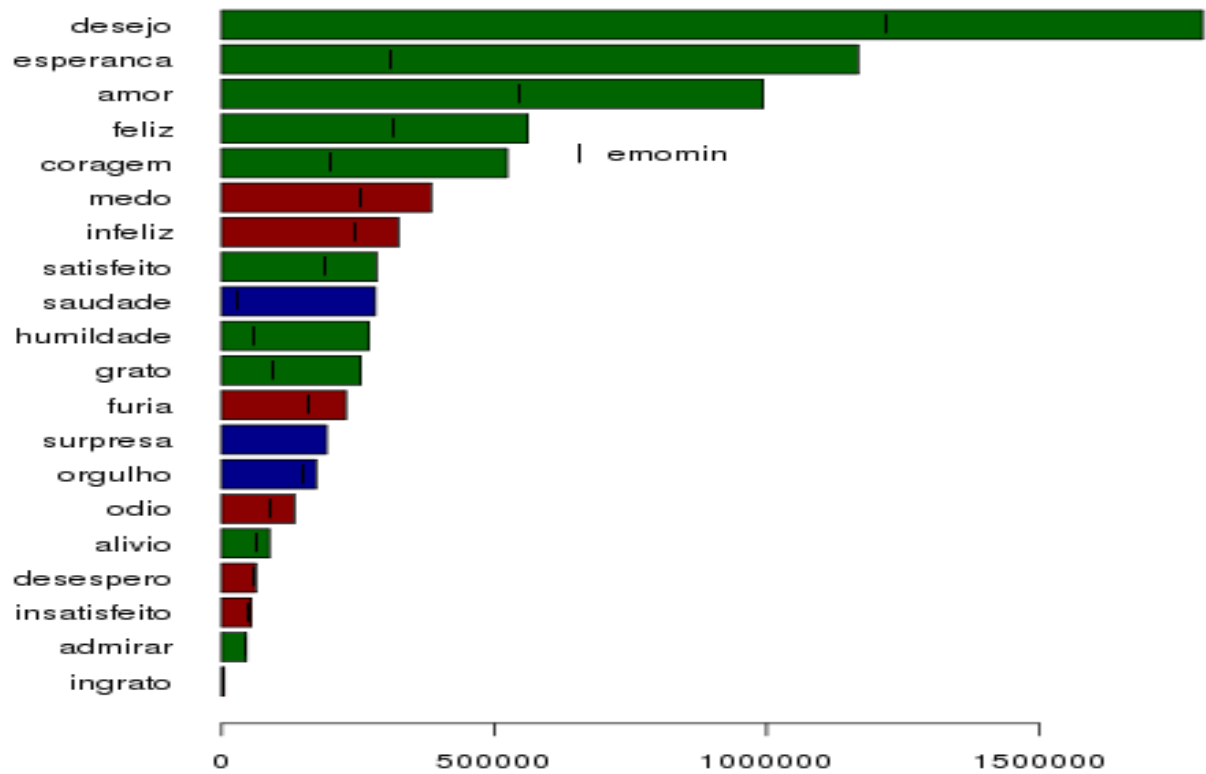
- monolingual corpora of Portuguese, compiled by most groups that work in Portuguese NLP (very few did not give us access), and some by us, amounting to ca. .5 billion words, <http://www.linguateca.pt/ACDC/>
- parallel corpora, compiled by Linguateca, some years after ENPC and acknowledging its influence, but also with significant differences: in aligned units, COMPARA (100,000), CorTrad (46,300), PoNTE (32,000) and PANTERA (13,000) <http://www.linguateca.pt/{COMPARA|CorTrad|PoNTE|PANTERA}/>

Very briefly, what emotions (19 groups)

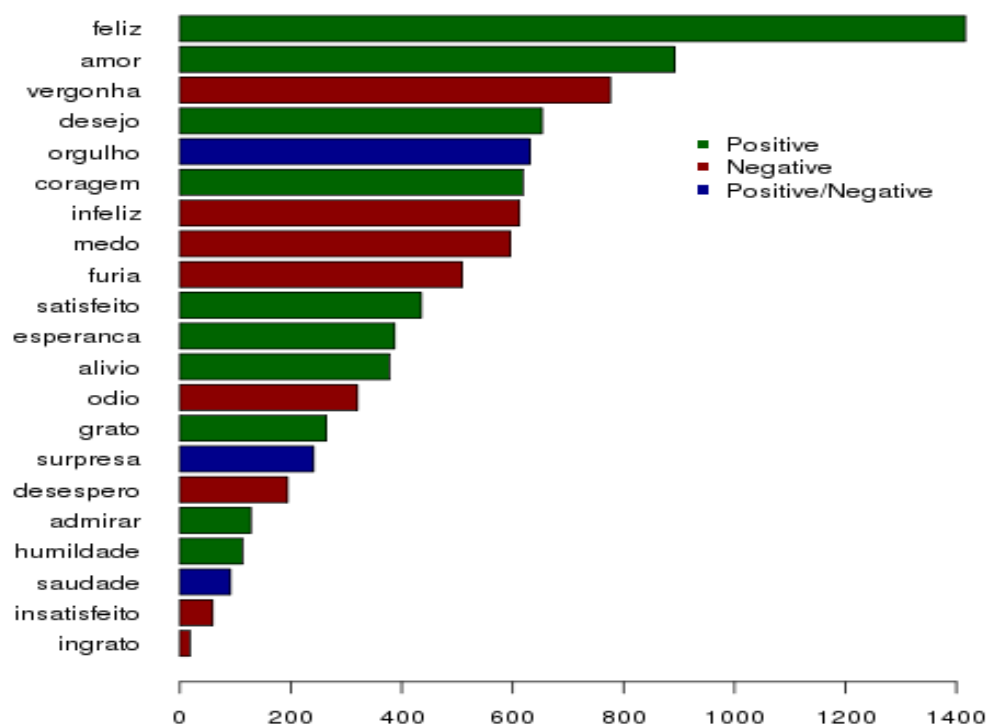
The English is just to give a pale idea. Note that, together with typical emotions, we have ethically-coloured emotions as well, for example “courage” or “ingratitude” (opposites of true emotions).

alívio	relief	admirar	admiration
amor	love	coragem	courage
desejo	desire	desespero	despair
esperança	hope	feliz	happy
fúria	anger	grato	grateful
infeliz	unhappy	humildade	humbleness
insatisfeito	insatisfaction	medo	fear
orgulho	pride	satisfeito	content
saudade	missing, nostalgia	surpresa	surprise
vergonha	shame		

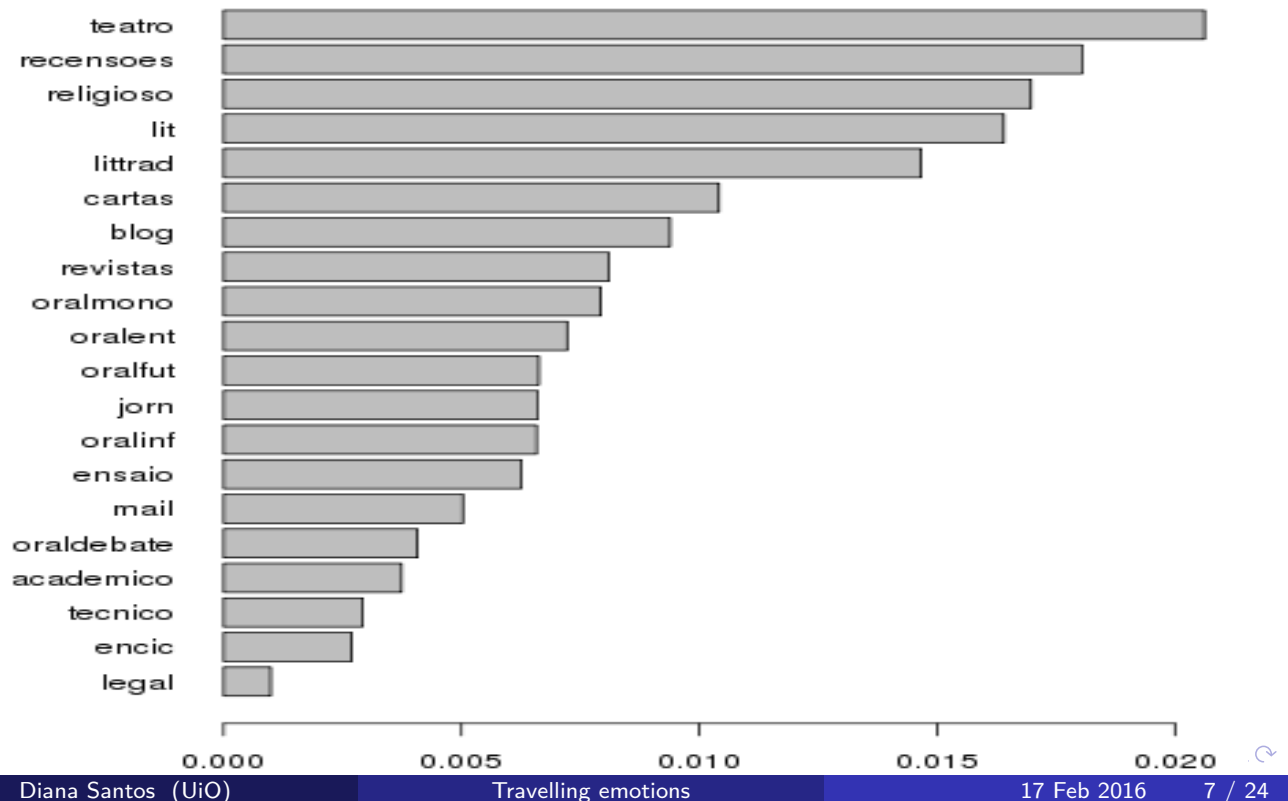
Emotion occurrences in the corpora, min and max



Lexical diversity: How many different words per emotion



Emotion occurrences per genre



Emotion infrastructure: examples og body and emotion

sorriso no rosto
felicidade estampada nos rostos
cabelos em fúria
coragem de dar as mãos
agitada, olhos arregalados
rancor irradiava-lhe dos olhos
de pé, terrível, pálido da raiva
sentir a adrenalina no sangue
não querem mais derramamentos de sangue
a melancolia da separação dera-lhe cabelos brancos
Agressivamente um dos vendedores resmungou entre dentes
mal seguro nas pernas, que vergavam sob o peso da ira

smile in the face
happiness written in the face
hair in anger
courage to give hands
frenetic, with wide-opened eyes
contempt emanated from his eyes
standing, awful, pale of hate
feel adrenalin in the blood
do not want no more blood sheddings
sadness of separation turned his hair white
aggressively mumbling with gritted teeth
hardly on his legs, falling with the hate weight

Cultural norms?

- My operational definition of culture is the ranking of values (and emotional values). While emotions can be universal, the way they are socially and linguistically apprehended and accepted or restrained is a major cultural difference, even among cultures so “similar” as Norwegian and Portuguese.
- Therefore, I am interested in finding evidence (empirical linguistic evidence) for different rankings involving the “same” emotions, as well as different attention to different parts of the emotional spectrum.

Some cultural differences

- Very generally, Germanic languages are focused on reality (facts, how the world is), while Romance languages are focused on moral values. Therefore there should be much more differentiation on “moral” emotions (those that are endowed with a judgment as well) in Portuguese than in Norwegian.
- Particular emotions have different (almost opposite) connotations in Portuguese and Norwegian. Example: *ciúmes* and *sjalu* (jealousy?)
- Particular emotions have different application areas/objects, such as *respeito* vs. *respekt* (respect?)
- Particular emotions are lexically connected in some languages and not in others: *admirar* can convey (both) surprise and respect, etc.

Where is the “travelling”?

- One instance of the “travelling” is in translation, and one (of the many) way(s) we can study the travelling of concepts is by looking at translation, in the particular cases when these words appear,
- Other ways could be by looking at the path these diachronic “words” have travelled (since proto-Indo-European ...) or at their contact through history (Norwegian from German from Latin...) (application at ILOS with Elizaveta Khachatryan)

In my case I am going to use direct (modern) translation between Portuguese and Norwegian as a tool, given the existence (or more accurately, the ongoing process of building) Norwegian-Portuguese parallel corpora.

PANTERA: translation between Norwegian and Portuguese

Portuguese And Norwegian Texts
for Education, Research and Acquisition of
relevant cultural and linguistic capabilities

- parallel
corpus modelled on COMPARA
but with more information
- an additional coupling to STIG (a
general system under development by
DMLF)
- growing every day: see quantitative
description at the PANTERA site,
[http:
//www.linguateca.pt/PANTERA/](http://www.linguateca.pt/PANTERA/)



The digitization of PANTERA texts has been financed by ILOS through several research assistants.

A word is better than a thousand pictures...



This was the best picture I could find in Google images.

Respeito in Portuguese

Using the AC/DC corpora (underlying Gramateca, 1.28 billion words), we can have a coarse picture of this word and other morphologically related

	CP	all
respeitar	20037	70375
respeito	18815 - 7336	200784
desrespeito	1561	9394
respeitável	1137	5554
desrespeitar	1124	7585
respeitado	963	8736
respeitoso	231	2917
respeitinho	34	41
desrespeitosamente	2	27
respeitosamente	0	925
	<hr/>	<hr/>
	44,014	306,338

Respeito in Portuguese: which emotion group?

A different but vexing question is: where do the words connected with respect appear in our emotion groups?

- some in ADMIRAÇÃO (admiration)
- some in HUMILDADE (humbleness)
- some in MEDO (fear)
- some in ORGULHO (pride)

Is this the proof that our emotion groups make no sense? Or that some emotions are connected, precisely through lexical means?

What does one *respeita* (or not) in Portuguese?

```
"respeito" [lema="por(\+.*)*"] @[pos="N.*"]
```

In CETEMPúblico: direito 727 princípio 101 lei 85 regra 79 vida 73 liberdade 69 norma 64 pessoa 59 diferença 49 valor 48 dignidade 47 ambiente 44 minoria 43 legalidade 42 tradição 36 constituição 34 ...

In all corpora: direito 997 vida 185 pessoa 161 diferença 152 lei 139 princípio 138 parte 121 regra 112 liberdade 109 dignidade 105 norma 84 valor 84 diversidade 78 trabalho 74 ...

What's the difference between *respeito por* and *respeito a*?

One way to look at the alignment (or not) of this “emotion” is to use translations as semantic data, something I have argued ever since I started my PhD.

So, using PANTERA and PoNTE, one can both

- look at the general picture,

```
([sema="emo.*(humildade|admirar|medo|orgulho).*" ] \&
[lema="respeitável|respeito|respeitar...respeitinho"])
ori    27   106196   2.54e-4
trad   11    75948   1.44e-4
```

- or look at specific cases (next slides)

Respeito in PANTERA: some Port-Nor examples

- *Carlos não manifestara qualquer **respeito**, amizade ou consideração pela sua pessoa* (Ger Alm11)
hadde Carlos aldri vist den ringeste **respekt**, hengivenhet eller omtanke for hans person
- *Tinha o rosto alegre, se bem bastasse um olhar seu para inspirar absoluto **respeito**.* (JorAma41)
Hun så alltid glad og munter ut, men hun visste å sette seg i **respekt** med et eneste blick.
- *a desagradável voz dele provavelmente não o impediria de tomar boa conta da quinta e ser **respeitado**.* (Pepete21)
den ubehagelige stemmen hans ville antagelig ikke være til hinder for at han kunne ta seg godt av gården og sette seg i **respekt**.
- *fez um gesto de **respeito** à mestra de noviças* (CamCB11)
Hun neide **respektfullt** overfor novisenes lærerinne

- Problemet er at jeg dessverre ikke kan **etterkomme** ønsket ditt. (AnnHol21)
*O problema é que infelizmente não posso **respeitar** a sua vontade.*
- Hvis politibetjenten foran ham skal kunne bli en kilde, er ikke arrogant og **respektløs** anbefalt oppførsel i journalistenes ABC. (ThoEng11)
*um comportamento arrogante e **desrespeitoso** não é a melhor estratégia.*
- Hvem er det som våger det hun gjør? Sende ungen sin alene over fjellene for å møte pappaen sin? Ikke mange. Det står det **respekt** av. (TorRen11)
*Poucos. Isso impõe **respeito**.*

Introduction (not conclusion!)

An emotion infra-structure to study emotions in translation and in language.

- Can we use this for anything interesting?
- Can we couple this with other annotation projects that deal with emotions? (technically, conceptually, methodologically...)
- Can we learn something about language, translation, emotions by querying these corpora?

I hope so!

