

Emotions in PANTERA

Diana Santos

Institute of Literature, Area Studies and European Languages, University of Oslo
d.s.m.santos@ilos.uio.no



22 May 2024



A small annotation project

- for the PANTERA corpus (Portuguese And Norwegian Texts for Education, Research and Acquisition of relevant knowledge...)
- to increase the information in the Norwegian part
- to allow for contrastive studies of the two languages
- to allow for translation studies

Main reference: Diana Santos. "PANTERA: a parallel corpus to study translation between Portuguese and Norwegian". *BELLS* 10.1 (2019).



Some data on PANTERA

Corpus in progress

- Created/available since 2013
- Currently 123 text pairs, 45,559 alignment units
- ca 700,000 words in Portuguese and 690,000 words in Norwegian
- Publicly available
- New pairs are (also) created by *kvalifiseringsoppgaver* in the POR4104 discipline

Follows the design of COMPARA (Frankenberg-Garcia & Santos, 2003), which was strongly inspired by ENPC (Johansson & Hofland, 1994). Main differences compared to the ENPC: public, varied choice of excerpts, richer annotation.

Emotion annotation

- NB! Different from sentiment analysis.
- We annotate reference to an emotion (PT *sentimento*, NO *følsel*), not emotion-loaded utterances.
- We believe that emotions are different in different languages (even if the words come from the same indo-European root)

Work for Portuguese: Diana Santos, Alberto Simões & Cristina Mota.
“Broad coverage emotion annotation”. *Language Resources and Evaluation* 56.4 (2022), pp. 857-879.

Examples of Portuguese emotions in Portuguese monolingual corpora

(The translation into English is merely indicative.)

Alívio	Relief
Admirar	Admiration
Amor	Love
Ausência	Absence
Coragem	Courage
Desejo	Desire
Desespero	Despair
Esperança	Hope
Feliz	Happiness
Fúria	Anger
Gen	Generic
Grato	Gratitude
Humildade	Humility
Infeliz	Unhappiness

Annotation for Norwegian, financed by the Texthub

- Marcin Włodek, trilingual and triple nationality, linguist, *hovedfag* in Portuguese.
- Lightly inspired by the Portuguese emotion groups, he should come up with suitable groups for Norwegian, and list words/expressions that belong in these groups.
- I should include this in the pipeline for Norwegian annotation, eventually also program disambiguation rules he came up with



Examples of Norwegian emotion groups

(The translation into English is merely indicative.)

annet	other
beundring	admiration
fravær	absence
frykt	fear
generisk	generic
glede	happiness
haap	hope
kjærlighet	love
lettelse	relief
medfølelse	compassion
mot	courage
misnøye	unsatisfaction
motvilje	hate
ønske	desire

Kinds of questions

- Which emotions are most frequent in ...
 - Norwegian
 - Norwegian compared with Portuguese
 - literature from Angola
 - non-fiction
 - author X
- How is a particular emotion translated into...
- What is the least translated emotion?
- What is the emotion with most different candidates?
- What is the emotion most mentioned about one self / the others?
- Are emotions most described in the past, present, or for the future?

- The Portuguese side is annotated with PALAVRAS, Bick (2014) (<https://edu.visl.dk/visl/pt/>)
- The Norwegian side is annotated with the Oslo-Bergen tagger, Johannessen et al. (2012) (<http://tekstlab.uio.no/obt-ny/english/>)
- Both corpora receive further annotation (and disambiguation) by programs written in VISLCG3 and in Perl, developed in Linguateca. They are publicly available from PANTERA's site, <https://www.linguateca.pt/PANTERA/>.

Which Norwegian authors use most emotion words?

Wassmo, Herbjørg	145	3294	0.0440
Nesbø, Jo	145	3641	0.0398
Kielland, Alexander Lange	248	6658	0.0372
Askildsen, Kjell	306	8442	0.0362
Vesaas, Tarjei	140	3910	0.0358
Haugen, Tormod	189	5340	0.0354
Renberg, Tore	456	13028	0.0350
Hansen, Erik Fosnes	120	3434	0.0349
Holt, Anne	1090	31587	0.0345
Loe, Erlend	99	2936	0.0337
Ibsen, Henrik	1414	42337	0.0334
Lindell, Unni	225	6761	0.0333
Hølmebakk, Sigbjørn	235	7252	0.0324
Fossum, Karin	750	24158	0.0310
Fosse, Jon	524	17070	0.0307
Gaarder, Jostein	1206	39589	0.0305

Translation between emo:frykt and emo:fear

No emo:frykt, addition of emo:medo by the Portuguese translation: 12 cases

Emo:medo, no emo:frykt in the Norwegian translation: 17 cases

- more common to speak about fear in Portuguese
- metaphorical use of *ameaçar* (threaten) common in Portuguese
- metaphorical use of *temer* (fear) as politeness more common in Portuguese
- stereotypical translation of macho males, with no fear? (Refsdal 2015)

Acknowledgements

- I thank ILOS for supporting financially the beginning of PANTERA
- I thank TextHub for financing this subproject
- I thank all people who corrected OCR and revised alignment throughout the years
- I thank FCCN for maintaining the servers where PANTERA is installed
- I thank Sigma for the computing storage and power in the cluster where I develop PANTERA

References

Bick, Eckhard. "PALAVRAS, a Constraint Grammar-based Parsing System for Portuguese". In Tony Berber Sardinha & Thelma de Lurdes São Bento Ferreira (eds.), *Working with Portuguese Corpora*, London/New York: Bloomsbury Academic, 2014, pp. 279-302.

Ana Frankenberg-Garcia & Diana Santos. "Introducing COMPARA, the Portuguese-English parallel translation corpus". In Federico Zanettin, Silvia Bernardini & Dominic Stewart (eds.), *Corpora in Translation Education*. Manchester : St. Jerome Publishing, 2003, pp. 71-87.

Johannessen, Janne Bondi, Kristin Hagen, André Lynam & Anders Nøklestad. "OBT+stat. A combined rule-based and statistical tagger". In Andersen, Gisle (ed.), *Exploring Newspaper Language. Corpus compilation and research based on the Norwegian Newspaper Corpus*. John Benjamins Publishing Company, 2012, pp. 51-65.

Johansson, Stig & Knut Hofland. "Towards an English-Norwegian parallel corpus". In U. Fries, G. Tottie & P. Schneider, *Creating and using English Language corpora, Papers from the fourteenth International Conference on English Language Research on Computerized Corpora, Zürich 1993*. Amsterdam & Atlanta: Rodopi, 1994, pp. 25-37.

Refsdal, Eva. "When "a girl" becomes "an attractive little number": Stereotyped representations of Latin America in literary translation and reception in 1960s Norway". PhD thesis, HF, University of Oslo, 2015.