

Paraphrasing Emotions in Portuguese



Cristina Mota
cmota@ist.utl.pt

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

Anabela Barreiro
anabela.barreiro@inesc-id.pt

NooJ conference, 5-7 June 2020

This work was supported by Portuguese national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020

Presentation



Introduction

Two projects cooperate

Emotions in language

Paraphrasing with eSPERTo

An empirical study

Procedure

The material

Quantitative data

First results

Encoding in NooJ

Future work



- ▶ **Emocionário** An attempt to organize emotions in Portuguese and annotate them in corpora
 - ▶ Mota & Santos (2015), Ramos et al. (2020),
<http://www.linguateca.pt/Emocionario/>



- ▶ **Emocionário** An attempt to organize emotions in Portuguese and annotate them in corpora
 - ▶ Mota & Santos (2015), Ramos et al. (2020),
<http://www.linguateca.pt/Emocionario/>
- ▶ **eSPERTo** A paraphrasing system using the NooJ linguistic engine, grammars, and lexicons
 - ▶ Barreiro, 2010; Mota et al., 2016; Barreiro et al., 2020
<https://esperto.hlt.inesc-id.pt/esperto/esperto/demo.pl>



Together, we might

- ▶ use the paraphrasing system for emotions
- ▶ use the emotions in the paraphrasing system
- ▶ evaluate the emotion classification so far
- ▶ identify interesting issues in the paraphrasing of emotions
- ▶ measure how many emotion paraphrases one can get from parallel text (currently, in two varieties of Portuguese)



Emocionário, a subproject of Linguateca, has as aims

- ▶ marking all **references to emotion** in text in our corpora (not expression of emotion!)
- ▶ classifying them in 24 not mutually-exclusive categories
- ▶ attempting to obtain guidelines both for emotion identifying and classifying
- ▶ getting a quantitative overview of emotions in Portuguese texts
- ▶ making the emotion annotation publicly available in the AC/DC corpora



- ▶ development and improvement of a paraphrasing system using linguistically enhanced resources developed within NooJ
 - ▶ provide suggestions of semantically identical expressions for text (re)writing
 - ▶ access to paraphrastic units that contribute to the linguistic and stylistic quality improvement of a text
 - ▶ language understanding when the use of certain expressions are unknown or unclear in an input text
 - ▶ show how vocabulary words are used when included in multiword expressions, thus significantly extending the usefulness of a simple dictionary
- ▶ integration of paraphrastic technology into other applications to make editing seamless



- ▶ eSPERTo's paraphrases have been tested in distinct scenarios
 - ▶ in a machine translation application [Barreiro, 2009]
 - ▶ in a dialogue system, to increase the linguistic knowledge of an intelligent virtual agent [Mota et al. 2016]
 - ▶ in a summarization and simplification system [Mota et al. 2016]
 - ▶ in a study of the differences in Portuguese varieties in literary texts [Barreiro and Mota 2018] [Rebelo and Barreiro 2018] [Barreiro et al., 2020]
 - ▶ in conversion of informal into formal language and vice-versa [Barreiro et al. 2018]
 - ▶ in transformation of technical versus non-expert expressions
 - ▶ in the editorial process [Barreiro et al., 2020]



Efforts have been undertaken to integrate and improve linguistic resources available for Portuguese, which were dispersed and lacked organic articulation among them; e.g. use of lexicon-grammar tables:

- ▶ paraphrases of human intransitive adjectives (HIA) [Mota et al. 2015]
- ▶ paraphrases of support verb constructions (SVC) with *fazer* [Mota et al., 2017]
- ▶ paraphrases of support verb constructions (SVC) with *ser de* [Mota et al, 2018]

Paraphrases of HIA and SVC with *fazer* and *ser de*



eSPERTo - System for Paraphrasing in Editing and Revision of Text

The screenshot displays the eSPERTo web interface. On the left, the 'Parameters' section includes options for Demo mode, Interface idiom (English), Resources idiom (Portuguese), Dictionary (PT-Dict_NEW), Sample text (SAN), and Paraphrasing options. Under Paraphrasing, 'Human intransitive adjective' and 'Predicate nouns with Vsup fazer' are checked. The 'Input file or text' section shows a text box with the sentence: 'Recentemente, a professora iraquiana foi de uma grande ousadia quando fez o protesto à porta da universidade.' Below this, a dropdown menu shows suggestions: 'houve na professora iraquiana' and 'teve'. The 'Results' section shows the original text with brackets around 'iraquiana' and 'foi de', and a dropdown menu with suggestions: 'houve na professora iraquiana', 'que é iraquiana', 'que é do Iraque', 'de origem iraquiana', 'de naturalidade iraquiana', 'de nacionalidade iraquiana', 'iraquiana', and 'do Iraque'. A 'Process results' button is visible at the bottom right of the interface.



The collaboration seemed to require at least the following steps

- ▶ automatically annotate parallel text with emotions
- ▶ revise it (to guarantee correctness and to assess the quality of the automatic annotation)
- ▶ identify cases where emotions had been maintained but with other words, and cases where different emotions had been chosen
- ▶ code them as paraphrastic resources

Also, identify whether the emotion groups could be used as sources of more paraphrases.



We had five different texts in two varieties of Portuguese (from Portugal and from Brazil), already sentence aligned in previous projects:

1. the independent translation of two works by David Lodge into the two varieties: *Therapy* (DL1) and *Changing Places* (DL3)
2. the independent translation of one work by Jostein Gaarder into the two varieties: *Hallo? Er det noen der?* (JG)
3. the independent translation of one work by Karl Ove Knausgaard into the two varieties: *Min Kamp 1* (KOK)
4. the adaptation of a book by Afonso Cruz, a Portuguese author, into Brazilian Portuguese: *Os livros que devoraram o meu pai* (AC)

Except for the last work, we only had excerpts.



alu = alignment units; rev. emo = number of emotions after human revision

text	words	alu	alu emo	emo	alu emo	rev. emo
AC _{PP}	19,214	1261	225	264	276	338
AC _{PB}	19,127		234	272	281	349
DL1 _{PP}	17,422	652	221	290	271	391
DL1 _{PB}	17,363		214	285	272	393
DL3 _{PP}	11,133	417	139	203	158	267
DL3 _{PB}	11,735		123	179	162	273
JG _{PP}	4,337	222	72	97	84	127
JG _{PB}	4,320		69	102	79	133
KOK _{PP}	7,212	310	58	84	68	111
KOK _{PB}	7,146		60	86	71	117
Total _{PP}	59,578	2862	715	938	857	1234
Total _{PB}	59,988		700	924	865	1265

Emotions per alignment unit



“Alignment unit” is what corresponds to one sentence in the original.
For the alignment units featuring at least one emotion:

Difference in number of emotions	number of alu
0	627
1	222
-1	77
2	8
-2	5
-3	2
3	1

there are 334 cases (out of 2862, thus 12%) where one translator referred to an emotion and the other did not!

Of the 627 alignment units with the same number of emotions, 596 had exactly the same groups of emotions, corresponding to 809 cases. In 280 al. units the emotion words were not the same.



A good source of inspiration for paraphrasing, after careful scrutiny.

- ▶ Surprisingly frequent addition or removal of emotions in the alignment units
- ▶ Fairly frequent change of emotion group

Examples

PP - só **teria confirmado os seus profundos receios**

PB - apenas **confirmado suas suspeitas mais profundas**

EN - merely confirmed his deepest **misgivings**

- ▶ Frequent change of the word used for the same group of emotions)

Examples

PP - isso não **teria surpreendido** MZ por aí além

PB - isto não **teria causado nenhuma surpresa** a MZ

EN - it would not have **surprised** MZ unduly

Multiword-level emotion



- ▶ We used Emocionário to annotate emotion at the **word level**, since it uses a dependency formalism
- ▶ But often emotions are expressed as multiwords and this had to be annotated manually

Examples

PP - Ela **fez beicinho** e disse: - É que eu tenho uma ideia fabulosa para uma série muito original, uma espécie de versão inglesa do Twin Peaks.

PB - Ela **fez um beicinho** e disse: - O negócio é que tive uma ideia sensacional para uma novela diferente, tipo Twin Peaks inglesa.

EN - She **pouted** a little and said, «It's just that I have this fabulous idea for an offbeat soap, a kind of English Twin Peaks.

"**fazer beicinho**" (EN: pout, pouting) is a MWU that expresses emotion of sadness, disappointment, disagreement.



Examples

pp - Arranquei a venda e pus-me aos saltos na sala, a **rogar** tantas **pragas** e a **dizer** tantos **palavrões** que acabei por ficar **chocado** e me calei.

pb - Arranquei a venda dos olhos e pulei de um lado para o outro ao redor da sala , **xingando** e **blasfemando** coisas tão terríveis que finalmente minha **vociferação** me fez silenciar

en - I tore off the blindfold and hopped round the room **cursing** and **blaspheming** so terribly I finally **shocked** myself into silence.

In Portuguese, the expressions "**rogar [] pragas**" and "**dizer [] palavrões**" are MWU that express emotion of anger, rage, fury.

Aligning MWU-level emotions with CLUE-Aligner 1



Browse Source Corpus
Browse Target Corpus
Browse Alignments

Help
About CLUE-Aligner

Load Corpora
2C
Save Alignments
Export Alignments

	Você	não	consegue	saber	de	nenhuma	sujeira	,	se	não	tiver	nada	para	contar	também	.
Quem																
gosta																
de																
falar																
desses																
assuntos																
é																
quem																
tem																
culpas																
no																
cartório																
.																

Quem gosta de falar desses assuntos | se não tiver nada para contar (S) Remove

tem culpas no cartório | saber de [] sujeira (S) Remove

	isto	não	teria	causado	nenhuma	surpresa	a	Morris	Zapp	,	apenas	confirmado	suas	suspeitas	mais	profundas
isso																
não																
teria																
surpreendido																
Morris																
Zapp																
por																
ai																
além																
,																
só																
teria																
confirmado																
os																
seus																
mais																
profundos																
receios																

teria surpreendido | teria causado [] surpresa a (S)
 profundos receios | suspeitas [] profundas (S)

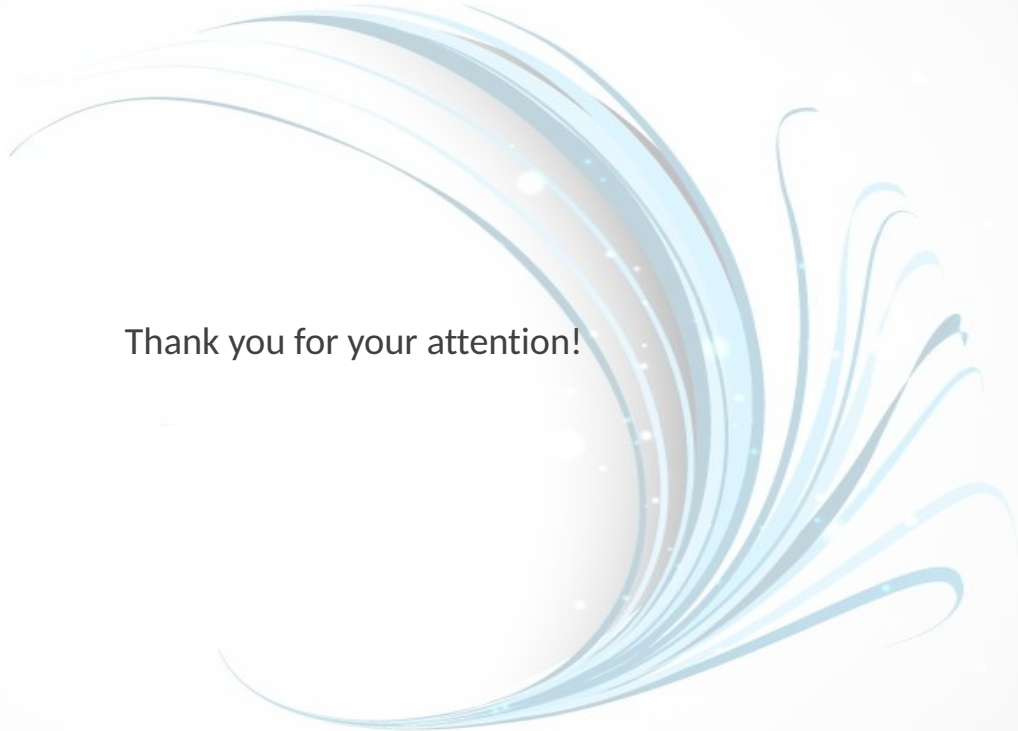
Cristina Mota, Diana Santos

How to include emotion paraphrases in NooJ

- ▶ NooJ resources and grammars underlie eSPERTo
- ▶ Examples that could be encoded
 - ▶ *falta de* +N(+Abstract) = *não ter* N(+Abstract)
 - ▶ *sentir falta de* N(+Human) = *ter saudade(s) de* N(+Human)
 - ▶ *com um sorriso* ADJ(*nervoso*) = *sorrindo* ADV(*nervosamente*)
 - ▶ N related to body parts or actions: *sorriso* (smile), *riso* (laugh), *olhos* (eyes) *choro* (cry), *ombros* (shoulders - shrug), *sobrolho* (eyebrows - frown), capitalizing from Esqueleto [Freitas et al. 2015]
 - ▶ opposite polarity emotions – *sorriso* (+) *triste* (-) (sad smile); *lastimável* (-) *sorte* (+) (unfortunate luck)
- ▶ (Due to Corona problems, we were not yet able to encode the grammars in NooJ, but this will be done as soon as possible)



- ▶ Analysis of MWU expressing emotion in large corpora (AC/DC)
- ▶ Design and construction of high coverage paraphrastic grammars in NooJ for emotion-related multiwords and other phrases expressing emotion, such as
 - ▶ **emotion-conveying idiomatic expressions**, which may correspond to a single word with identical meaning, e.g. *ter pena de* (have pity for), *sentir falta de* (miss), *encolher os ombros* (shrug shoulders), *fazer beicinho* (pout), *(falar) com o coração nas mãos* (heart in [] hands), *de cabelos em pé* (hair raising)
 - ▶ **emotion quantification**, e.g. *gostar muito*, paraphrase of *amar* (like a lot = love); *não gostar nada*, paraphrase of *detestar* (do not like at all = hate), also using **opposite emotions**
 - ▶ **combination of different emotions**, e.g. *sorriso triste* (sad smile); *má sorte* (bad luck).
- ▶ Evaluation of emotional paraphrases with users



Thank you for your attention!