

# Colouring COMPARA

contrastive and monolingual  
colour studies in  
English and Portuguese

Rosário Silva  
Susana Inácio  
Diana Santos

COMPARA  
www.linguateca.pt/COMPARA/

- the largest edited parallel corpus (publicly available since 2000)
  - 3 million words, 100 000 "aligned sentences"
- 75 bi-texts (original and translated literature) in Portuguese and English, 35 authors and 45 translators
- developed at Linguateca
  - (project team: Ana Frankenberg-Garcia, Diana Santos, Pedro Sousa, Rosário Silva and Susana Inácio)
- special features: manually revised texts, alignment and grammatical annotation

Colour seen from the viewpoint of:

- Language
  - Portuguese
  - English
- Author
- Lexico-grammatical patterns
- Translation

Why COLOUR?

- We wanted to start the semantic annotation of COMPARA
- An easy to grasp and relatively-contained lexical category
- The subject of many arguments in linguistics
- Its use is frequent in language, both literally and metaphorically

Colouring COMPARA – Criteria

- Straightforward colour words, regardless of word class (red, blue, verde, cor-de-rosa,...)
- Words that convey colour, but are not in themselves a colour (blond, tanned, corado,...)
- Words where colour is used metaphorically (greyhound, blackmail, sorriso amarelo, correio azul...)
- The word "colo(u)r" itself (colo(u)red, colo(u)rful,...)

Classification of words marked for colour

- <colour>
  - sub-category
    - <Beige>
    - <Black>
    - <Blue>
    - <Brown>
    - <Gold>
    - <Green>
    - <Grey>
    - <Orange>
    - <Pink>
  - <Purple>
  - <Red>
  - <Silver>
  - <White>
  - <Yellow>
  - <Other>
  - <Multiple>
  - <Unspecified>
- <colour:race>
- <colour:human>
- <colour:wine>
- <colour:original>

7

## Mark-up examples


**Example of <colour>**  
The old one had a beard, <colour><Blue> bluish </Blue></colour> lips...

**Example of <colour:race>**  
Either yours or some other <colour:race> white </colour:race> man's.

**Example of <colour:human>**  
Marjorie <colour:human> blushes </colour:human>.

**Example of <colour:wine>**  
There was imported <colour:wine> white </colour:wine> wine and we had oysters.

**Example of <colour:original>**  
The <colour:original> black </colour:original> sheep of the family.


Colouring COMPARA 

8 Colouring COMPARA

## Colour frequency in COMPARA 10.0

**Portuguese:**  
4609 tokens, 409 types, 209 lemmas

**English:**  
4548 tokens, 427 types (not yet lemmatised)

Colouring COMPARA 

9


## Which language uses more colour?

**COLOUR DENSITY** (tokens per 1000 words)

- English: 3.0
- Portuguese: 3.3

**COLOUR DIVERSITY** (types per 1000 words)

- English: 12.6
- Portuguese: 8.0

Colouring COMPARA 

10

## Which language uses more...?

**COLOUR:RACE** (per 10 000 words)

- English: 5.0
- Portuguese: 5.0

**COLOUR:HUMAN** (per 10 000 words)


- English: 2.0
- Portuguese: 2.0

**COLOUR:WINE** (per 10 000 words)

- English: 0.1
- Portuguese: 0.1

**COLOUR:ORIGINAL** (per 10 000 words)


- English: 2.1
- Portuguese: 1.4

Colouring COMPARA 

11

## Colour per English-speaking author


English-speaking authors (words)	Colour types	Colour tokens	Freq./1000 words
Barnes, Julian (59965)	46	121	2.02
Carrol, Lewis (11089)	7	11	0.99
Conrad, Joseph (12162)	19	41	3.37
Gordimer, Nadine (88027)	92	617	7.01
Heller, Joseph (13252)	12	25	1.89
Ishiguro, Kazuo (84847)	23	69	0.81
James, Henry (47026)	22	65	1.38
Lodge, David (184377)	95	534	2.90
McEwan, Ian (41781)	44	103	2.47
Poe, Edgar Allan (24411)	7	20	0.82
Shelley, Mary (20876)	6	12	0.57
Trollope, Joanna (84363)	101	312	3.70
Wilde, Oscar (25647)	52	166	6.47
Zimler, Richard (61195)	63	212	3.46

Colouring COMPARA 

12

## Colour per Portuguese-speaking author

Portuguese-speaking authors (words)	Colour types	Colour tokens	Freq./1000 words	Portuguese-speaking authors (words)	Colour types	Colour tokens	Freq./1000 words
A. Azevedo (39789)	33	159	4.00	Lidia Jorge (17363)	20	76	4.38
A. Dourado (40910)	51	364.5	8.91	M. Assis (79987)	21	72	0.90
C. C. Branco (15679)	5	8	0.51	M. Almeida (11271)	8	13	1.15
C. Buarque (20857)	31	112.5	5.39	Marcos Rey (18465)	19	53	2.87
Eça Queirós (36716)	36	255.5	6.96	M. Carvalho (20837)	19	37	1.78
Jô Soares (18526)	13	43	2.32	M.Sá-Carneiro (21853)	41	190	8.69
Jorge Sena (42476)	23	52	1.22	Mia Couto (16002)	12	46	2.87
José C. Pires (14892)	17	34	2.28	Osman Lins (17805)	30	103.5	5.81
José Alencar (3197)	10	33	0.32	Patrícia Melo (31282)	27	112.5	3.60
José Aguaiusa (4749)	9	19	4.00	Paulo Coelho (41521)	13	29	0.70
J. Saramago (63762)	24	123.5	1.94	R. Fonseca (58494)	32	174	2.97

Colouring COMPARA 



19

## The translation of colour

PT>EN example: translation of <Amarelo> in Eça de Queirós

	Amarelo
Eça de Queirós	15
Translator (EQ)	12
	-3

	Amarelo	Eça de Queirós	Translator (EQ)
PPEQ1(3):	amarelo	yellow	
PPEQ1(32):	amarelado	yellowing	
PPEQ1(43):	amarela	yellow	
PPEQ1(126):	amarela	yellow	
PPEQ1(272):	amarela	yellow	
PPEQ1(276):	amarelo	yellow	
PPEQ2(291):	amarela	sallow	
PPEQ2(824):	amarela	yellow	
PPEQ2(1092):	cor de milho	corn-coloured	
PPEQ2(1184):	amarelo	yellow	
PPEQ2(1320):	amarelo	yellow	
PPEQ3(7):	amarelado	pale	
PPEQ3(33):	amarelenta	pale	
PPEQ3(71):	amarelada	wan	
PPEQ3(142):	amarelo	yellow	

Translations do not fall into any colour categories

Colouring COMPARA

20

Colouring COMPARA

## Colour seen from the viewpoint of:

- Language
  - Portuguese
  - English
- Author
- Lexico-grammatical patterns
- Translation

Colouring COMPARA

21

Colouring COMPARA

## Many more possibilities:

- Colour in different texts
- Colour in different varieties of Portuguese or English
- Diachronic studies (1818...2002)
- Colour in translated and non-translated texts

Colouring COMPARA

22

Colouring COMPARA

## Besides colour


- Semi-automatic annotation of other semantic fields possible
- Interested researchers contact us at:
   
[www.linguateca.pt/COMPARA/](http://www.linguateca.pt/COMPARA/)

Colouring COMPARA


23


## Acknowledgment

- This work was done in the scope of Linguateca. contract no. 339/1.3/C/NAC. project jointly funded by the Portuguese Government and the European Union.



Programa Operacional Sociedade do Conhecimento





Fundação para a Computação Científica Nacional

Colouring COMPARA