


Colouring COMPARA

contrastive and monolingual
colour studies in
English and Portuguese

Rosário Silva
Susana Inácio
Diana Santos




- COMPARA
- Colouring COMPARA
- English-speaking authors
- Contrastive analysis (Portuguese vs. English)
- Morphosyntactic patterns in Portuguese and in English
- Originals vs. Translations

Colouring COMPARA 


COMPARA

- An English-Portuguese publicly available parallel corpus (since 2001)
- developed at Linguateca, project leaders: Ana Frankenberg-Garcia & Diana Santos
- manually revised texts, alignment and grammatical annotation (from PALAVRAS and CLAWS, in progress)
- several innovative features

Colouring COMPARA 

COMPARA contents

- 75 pairs, corresponding to 72 fiction works
- 1.436 million words in Portuguese (Africa, Brazil and Portugal)
- 1.549 million words in English (Great-Britain, America and South Africa)
- 97,728 alignment units
- see <http://www.linguateca.pt/COMPARA/>

Colouring COMPARA 

Colour research by:


- Language
 - Portuguese
 - English
- Language variety
 - Portuguese (Portugal, Brazil, Mozambique and Angola)
 - English (United Kingdom, United States and South Africa)

Colouring COMPARA 

Colour research by:

- Text
- Author
- Translated or Original text(s)
- Date of publication

Several outputs available (concordances and distributions)

Colouring COMPARA 

7 Colouring COMPARA

Why COLOUR?

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

Colouring COMPARA

8 Colouring COMPARA

What is COLOUR? – Criteria

- All straightforward words conveying colour, regardless of word class
- Words that convey colour but are not in themselves a colour
- Words that have colours in them but have gone beyond the mere colour reference
- The word “colour” and its derivations.

For more information see <http://www.linguistica.pt/COMPARA/Doc/AnotacaoCorCOMPARA.pdf> (continually updated)

Colouring COMPARA

9 Colouring COMPARA

COLOUR categories and groups

- Colour →
 - 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

Colouring COMPARA

10

Colour categories - examples

colour	amber-coloured; blackened; lime-green; red-and-white; watery-blue; white; yellowing...
colour:race	whites; blacks; white-skinned; coloureds...
colour:human	blond; blushed; brunette; red-haired; rosy-cheeked; suntanned; white-bearded...
colour:wine	white; red.
colour:original	blacklist; bluebirds; brownies; whitewash; eggwhite...

Colouring COMPARA

11 Colouring COMPARA

Quantitative data

- Instances marked as colour in COMPARA: 4,609 in PT and 4,548 in EN
 - 0.0032%
 - 0.0029%
- Number of different colours in COMPARA: 409 in PT (209 lemmas) and 427 in EN

Colouring COMPARA

12 English-speaking authors

ENGLISH-SPEAKING AUTHORS IN COMPARA

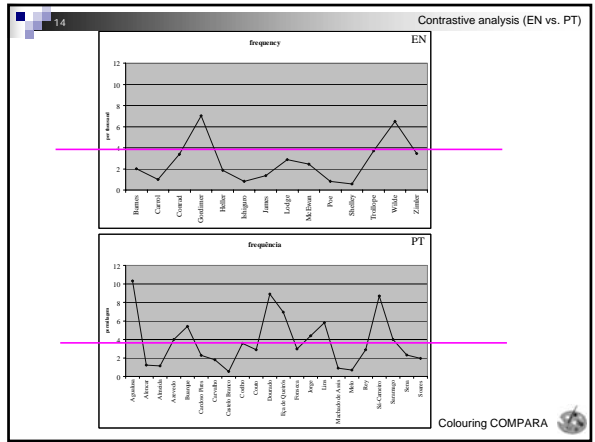
- Julian Barnes
- Lewis Carrol
- Joseph Conrad
- Nadine Gordimer
- Joseph Heller
- Kazuo Ishiguro
- Henry James
- David Lodge
- Ian McEwan
- Mary Shelley
- Joanna Trollope
- Oscar Wilde
- Richard Zimler

Colouring COMPARA


13 English-speaking authors

English Authors	Texts and Dates	# types	# words	freq.
Barnes	EBJH1 (1984) e EBJH2 (1989) e EBJH3 (1991)	46	121	2,02
Carrol	EBLC1 (1871)	7	11	0,99
Conrad	EBJC1 (1902)	19	41	3,37
Gordimer	ESNG1 (1990) e ESNG2 (1979) e ESNG3 (1981) e ESNG4 (1974)	92	617	7,01
Heller	EJHH (1990)	12	25	1,89
Ishiguro	EBKH1 (1995) e EBK12 (1989)	23	69	0,81
James	EUH1 (1880) e EUH2 (1895) e EUH3 (1893)	22	65	1,38
Lodge	EBDL1 (1995) e EBDL2 (1989) e EBDL3 (1975) e EBDL4 (1980) e EBDL5 (1991) e EBDL6 (1984)	95	534	2,90
McEwan	EBIM1 (1992) e EBIM2 (1998) e EBIM3 (1975)	44	103	2,47
Poe	EUEP1 (1837)	7	20	0,82
Shelley	EBMS1 (1818)	6	12	0,57
Trollope	EBTF1 (1996) e EBTF2 (1993) e EBTF3 (1995)	101	312	3,70
Wilde	EBOW1 (1890)	52	166	6,47
Zimler	EURZ1 (1998) e EURZ2 (2000)	63	212	3,46

Colouring COMPARA




15 Most colourful authors (no. of different colour types)



102
*blue, white, black, red,...
scarlet, hazel, peacock-blue, saffron-yellow, purplish, red-and-ochre, emerald-green...*


Joanna Trollope



51
preto, branco, negro, verde,... esbranquiçado, pratear, verde-garrafa, azuleleca, multicolor, rubro, alvaído,....


Aurran Dourado

Least colourful authors (no. of different colour types)



6
black, blue, green, white, red, blackness.

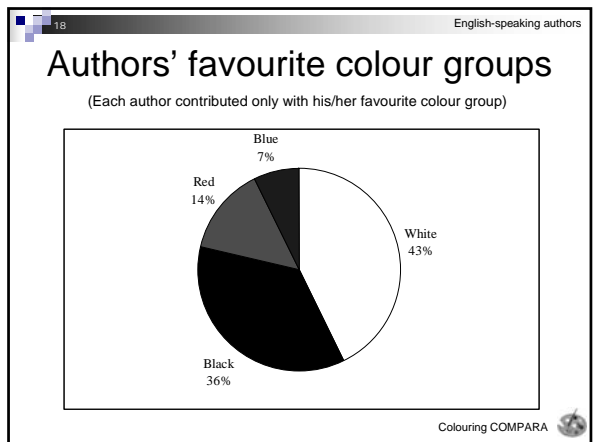
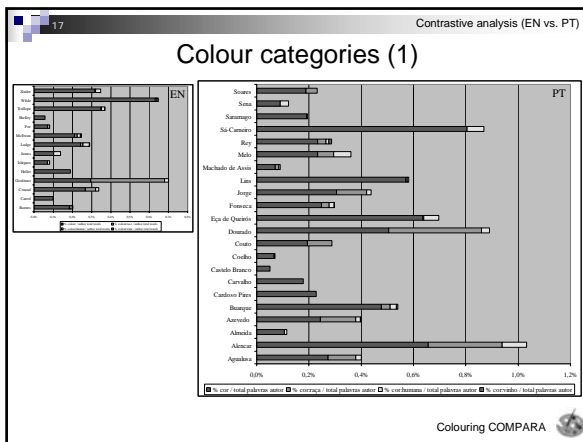
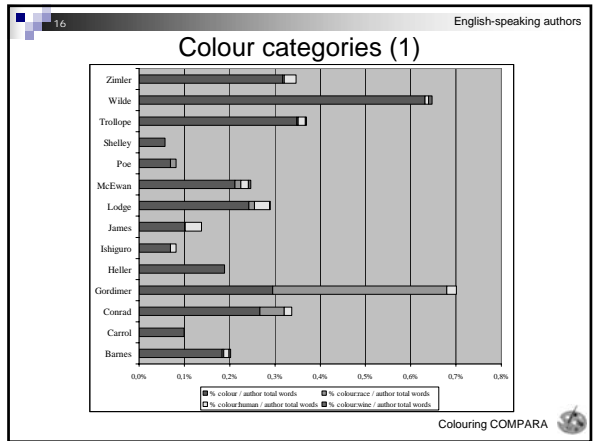
Mary Shelley

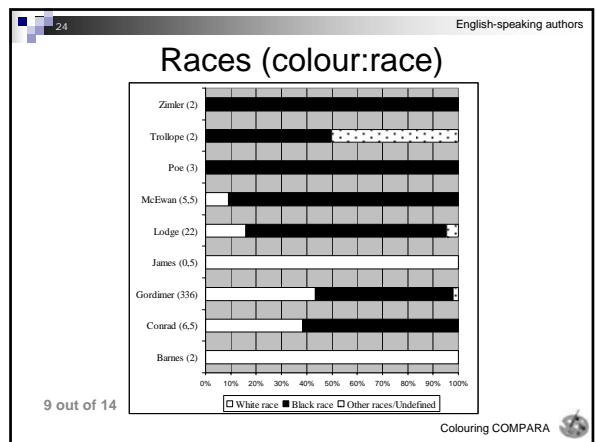
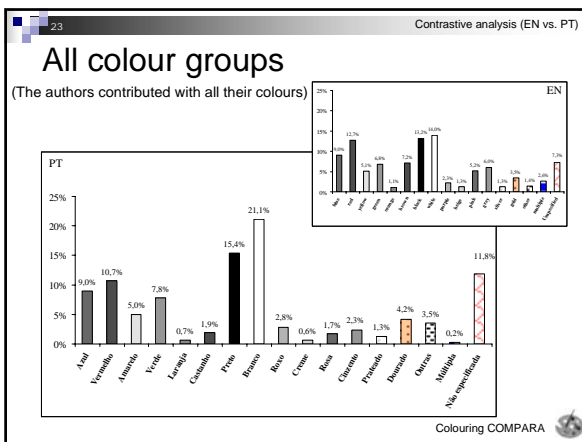
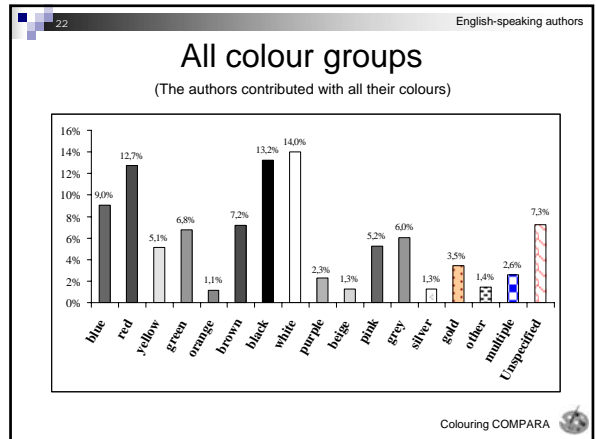
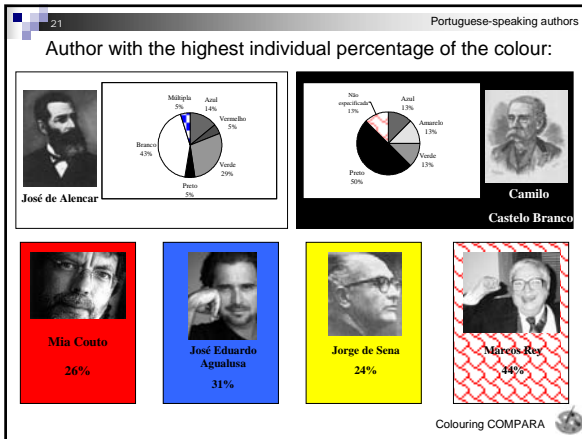
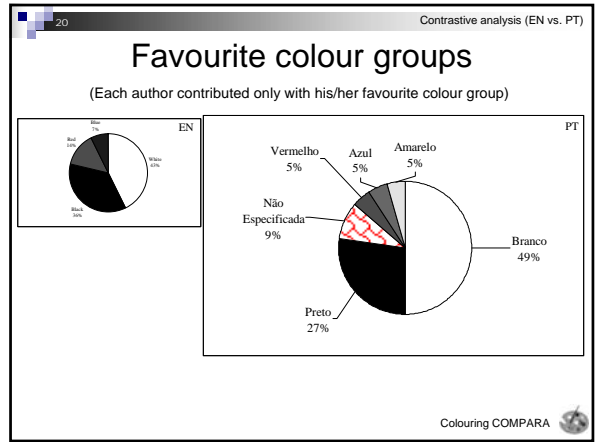
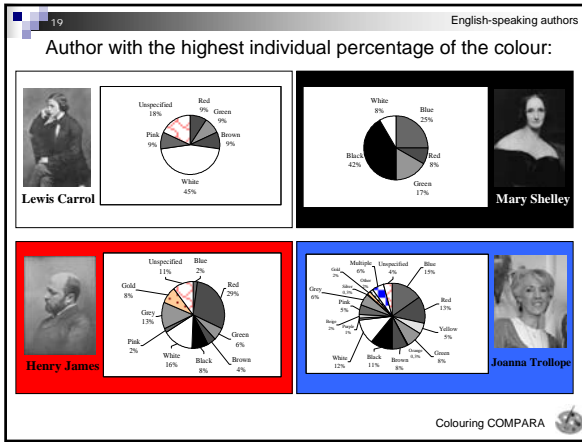


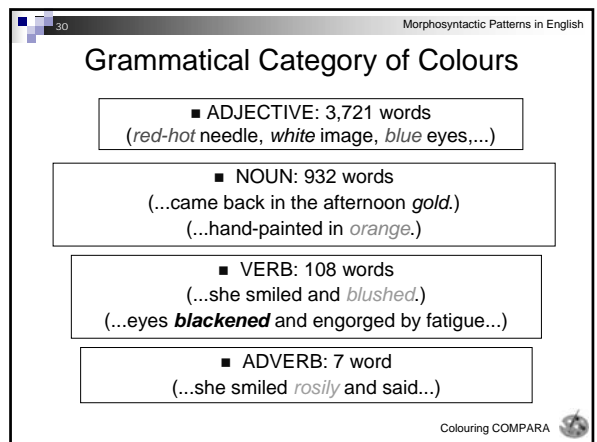
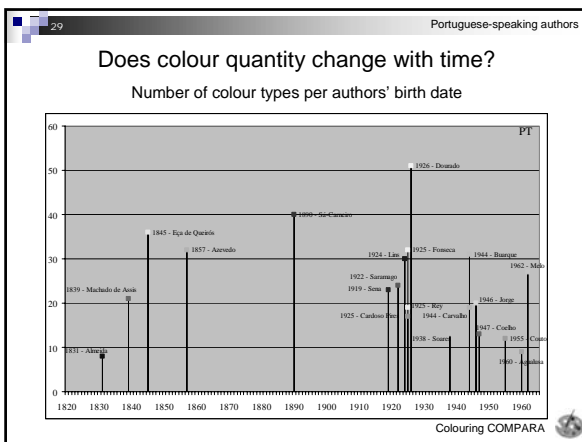
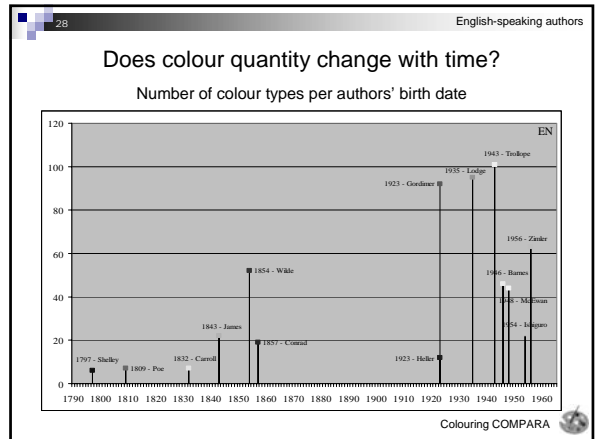
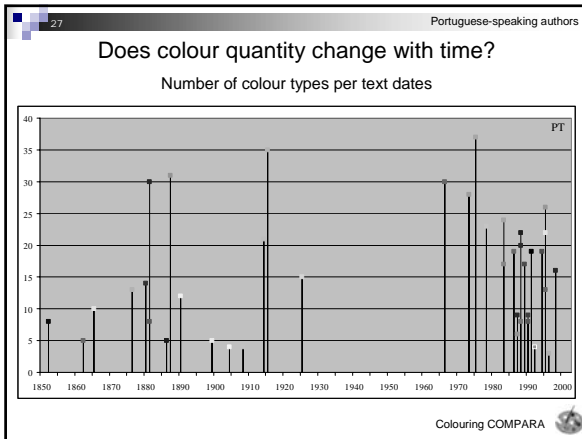
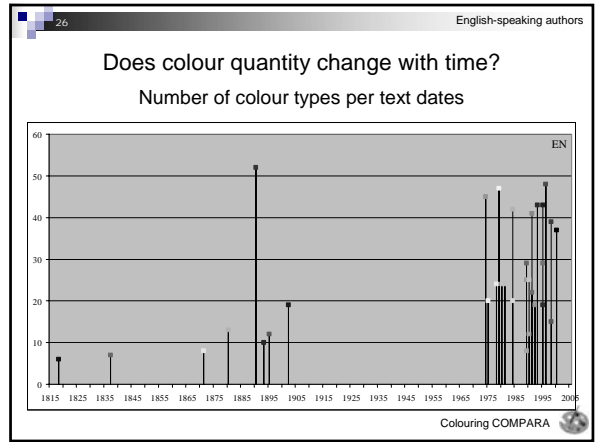
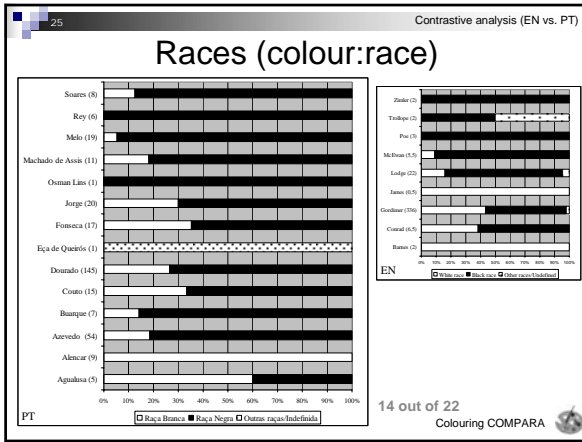
6
negro, verde, amarelo, azul, cor, preto.

Castelo Branco

Colouring COMPARA







31 Syntactic Patterns in Portuguese

Grammatical Category of Colours

- ADJECTIVE: 3,193 words (161 lemmas)
(carro *verde*, fato *azul*)
- NOUN: 1,349 words (93 lemmas)
(...o *verde* era uma cor que..., ...estou em pleno *azul*, suspenso...)
- VERB: 107 words (15 lemmas)
(...muito além daquela serra, que ainda *azula* no horizonte...)
(...as saladas *esverdinham* à hora do jantar.)
- ADVERB: 1 word (1 lemma)
(*purpureamente*)
- ADJn: 29 words (16 lemmas)
(-- No entanto gosto do *azul* -- prosseguiu, a cliente -- e o *verde* é muito bonito.)
- ADJ_N: 27 words (5 lemmas)
(Mas Isidoro, mesmo *preto*, não era capaz de uma coisa dessas.)

Colouring COMPARA

32 Morphosyntactic Patterns in English

Morphosyntactic Patterns using Colour

COLOUR + NOUN (1,138)

- hair 49 (*brown* hair, *reddish* hair, *wine-red* hair...)
- eye 45 (*red* eye, *watery-blue* eyes,...)
- silk 22 (*crimson* silk, *amber-coloured* silk,...)
- paper 16 (*yellowing* paper, *brown* paper,...)
- light 16 (*purple* light, *greyish-blue* light,...)
- velvet 11 (**black** velvet, *purple* velvet,...)
- sky 11 (*indigo* sky, *grey* sky,...)
- plastic 11 (*pale-grey* plastic, *white* plastic,...)
- bear 10 (*brown* bear, **black** bear,...)

Data from EN originals only, without colour.original, colour.race, colour.human and colour.wine

Colouring COMPARA

33 Morphosyntactic Patterns in Portuguese

Morphosyntactic Patterns using Colour

NOUN + COLOUR (914)

- cabelo 41 (cabelos *brancos*, cabelo **negro**,...)
- olho 32 (olhos *azuis*, olhos *castanhos*, olhos *verdes*,...)
- luz 21 (luz *vermelha*, luz *anilada*,...)
- seda 20 (seda *verde*, seda *roxa*,...)
- papel 19 (papel *colorido*, papel *amarelo*,...)
- mármore 13 (mármore **negro**, mármore *verde*,...)
- vestido 10 (vestido *azul*, vestido *prateado*,...)
- céu 9 (céu *azul*, céu *cinzento*, céu **negro**,...)

Data from PT originals only, without cor.original, cor.race, cor.humana and cor.vinho

Colouring COMPARA

34 Morphosyntactic Patterns in Portuguese

Morphosyntactic Patterns using Colour

COLOUR + NOUN (45)

- mar 6 (*verdes* mares...)
- praia 2 (*alvas* praias...)
- areia 2 (*brancas* areias...)
- cor 2 (*alva* cor...)
- rede 2 (*alva* rede...)
- fio 2 (*alvos* fios...)
- pelúcia 2 (*verde* pelúcia...)

Data from PT originals only, without cor.original, cor.race, cor.humana and cor.vinho

Colouring COMPARA

35 Morphosyntactic Patterns in English

Morphosyntactic Patterns using Colour

COLOUR + "and" + COLOUR (98)

- pink and (colour) 17
- black** and (colour) 10
- red and (colour) 9
- green and (colour) 7
- white and (colour) 6
- orange and (colour) 6
- blue and (colour) 6

Data from EN originals only

Colouring COMPARA

36 Morphosyntactic Patterns in Portuguese

Morphosyntactic Patterns using Colour

COLOUR + "e" + COLOUR (32)

- branco e (colour) 7
- azul e (colour) 5
- preto** e (colour) 3
- ouro e (colour) 2
- verde e (colour) 2
- vermelho e (colour) 2
- amarelo* e (colour) 2

Data from PT originals only

Colouring COMPARA

37 Morphosyntactic Patterns in English

Morphosyntactic Patterns using Colour

COLOUR (hyphenated words)

(396) Colour can appear on both sides of the hyphen

- colour-[borrowed from something] 85 (*cherry-red*,...)
- colour-[hue modifier] 66 (*deep-blue*,...)
- colour-[applied to something] 66 (*green-shirted*,...)
- colour-["colo(u)red"] 61 (*carrot-colored*,...)
- colour-[colour] 46 (*pink-brown*,...)
- colour-[hair-related] 44 (*red-haired*,...)
- colour-[skin-related] 20 (*olive-skinned*,...)
- colour-[eye-related] 8 (*gray-eyed*,...)

EN-side of the corpus Colouring COMPARA

38 Morphosyntactic Patterns in Portuguese

Morphosyntactic Patterns using Colour

COLOUR (hyphenated words)

(219)

- [hue modification] 104 (*verde-vivo*, *vermelho-escuro*,...)
- cor-de-... 62 (*cor-de-rosa*, *cor-de-rato*,...)
- [borrowed from something] 46 (*verde-musgo*,...)
- [several colours] 6 (*verde-azul*, *rosa-e-azul*,...)
- [applied to something] 1 (*saliva-laranja*)

PT-side of the corpus Hyphenated form of clitics in PT not included Colouring COMPARA

39 Morphosyntactic Patterns in Portuguese

Morphosyntactic Patterns using Colour

COR + DE + ...

(105)

- laranja 12
- azeitona 8
- areia 7
- cobre 4
- prata 4

PT-side of the corpus Colouring COMPARA

40

Colour - Originals vs. Translations

Term	Value
T (JB)	0
T (LC)	8
T (RC)	-6
T (NG)	-22
T (HH)	0
T (KL)	-6
T (HL)	-14
T (DL)	141
T (MO)	-5
T (EP)	1
T (MS)	-6
T (JT)	-22
T (OW)	-26
T (RZ)	-7

Colouring COMPARA

41 Originals vs. Translations - EN texts

Colour categories - Originals vs. Translations

Colouring COMPARA

42 Originals vs. Translations

Example: Julian Barnes

	colour:human	colour:human	Julian Barnes	Translator (JB)
Julian Barnes	7	EBJB1(228):	white-haired	branco
Translator (JB)	4	EBJB1(240):	blonder	louro
	-3	EBJB2(613):	blond	louras
		EBJB3(494):	blond	louro
		EBJB3(865):	tan	bronzeados
		EBJB3(881):	silver-haired	∅
		EBJB3(888):	blushes	∅

No translation

Translations fall into other colour categories

Colouring COMPARA

43 Originals vs. Translations

Example: Julian Barnes

	Red
Julian Barnes	16
Translator (JB)	9
	-7

No translation

Translations fall into other colour categories

Red	Julian Barnes	Translator (JB)
EBJB1(363):	red	vermelhas
EBJB1(364):	red	vermelhas
EBJB1(490):	red	∅
EBJB1(913):	reddish	avermelhados
EBJB1(929):	reddish	avermelhado
EBJB1(936):	reddish	avermelhado
EBJB2(448):	red	ruivo
EBJB2(449):	red	ruivo
EBJB2(457):	red	ruivos
EBJB2(461):	red	ruiva
EBJB2(616):	reddish	vermelhusca
EBJB2(717):	red	vermelho
EBJB2(774):	red	vermelhas
EBJB3(169):	burgundy	cor-de-vinho
EBJB3(290):	encrimsoned	corei
EBJB3(709):	red	vermelha

Colouring COMPARA

44

Which language uses more colour?

COLOUR DENSITY (per 1,000 words)

- English: 3.041
- Portuguese: 3.315

COLOUR DIVERSITY (per 1,000 words)

- English: 12.569
- Portuguese: 7.980

without colour:original

Colouring COMPARA

45

Which language uses more...?

COLOUR:RACE (per 10,000 words)

- English: 5.026
- Portuguese: 4.997

COLOUR:HUMAN (per 10,000 words)

- English: 2.055
- Portuguese: 1.941

COLOUR:WINE (per 10,000 words)

- English: 0.145
- Portuguese: 0.079

COLOUR:ORIGINAL (per 10,000 words)

- English: 2.055
- Portuguese: 1.414

Colouring COMPARA

46

We presented a large repository of aligned coloured texts available to every researcher interested in Portuguese or English.

COMPARA is also open to the semi-automatic annotation of other semantic fields, in collaboration with interested researchers who volunteer to revise the final annotation.

Colouring COMPARA

47

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.

POS_CONHECIMENTO
Programa Operacional Sociedade do Conhecimento

União Europeia
FEDERIFE

FCCN
Fundação para a Ciência e a Tecnologia

Colouring COMPARA