

Colours, clothing and food in CorTrad: why corpus-based translation studies are revealing

Diana Santos (Linguatca & Univ. of Oslo)

Stella E. O. Tagnin (DLM-USP)

Elisa Duarte Teixeira (Projeto CoMET)

- ICAME 2011 (Oslo, 1-5 June 2011) -

Table of contents

- Short introduction to CorTrad
- Description of the current semantic information and its human revision
 - Size and distribution
 - Annotation rationale
- Some contrastive issues

Parallel corpora and translation

- ✓ Isabelle (1992), McEnery & Wilson (1993) and Somers (1993) – potential for **machine translation**
- ✓ Johansson & Hofland (1994), Santos (1995) and Tagnin (2004) – potential for **contrastive studies**
- ✓ Malmkjaer (1998), Frankenberg-Garcia (1999) and Peters *et al.* (2000) – potential for **translation teaching**

CorTrad in a nutshell

- ✓ Annotated, multiversion pt-en-pt parallel corpus
- ✓ Started in May 2008
- ✓ Partnership:
 - ✓ Linguateca (design, development & implementation of computational framework)
 - ✓ CoMET Project (design & text collection and edition)
 - ✓ NILC - Inter Institutional Center for Computational Linguistics (web hosting)

CorTrad in a nutshell

- ✓ **Innovations** compared to other parallel corpora:
 - ✓ **Multiversion** format allows
 - ✓ comparison of different translation stages
 - ✓ study of revision process
 - ✓ **Refined search system** – tailored especially for each genre and text type
 - ✓ **Semantic information** – added and human-revised

CorTrad Parallel Subcorpora

Journalistic

Scientific
(pt → en)
1,076 texts

Technical-
Scientific

Cookbook
(pt → en)
130,000 words

USP PhD
abstracts
(pt → en)
Coming soon!

Literary

Australian
Short Stories
(en → pt)
28 texts

Canadian
Short Stories
(en → pt)
Coming soon!

Alice in
wonderland
(en → pt)
Coming soon!

Legal

Mercosul
Agreements
(pt → ← en)
Coming soon!

Journalistic (Science): Revista FAPESP



Original
(Brazilian Portuguese)



Published translation
(online publication)



Technical-Scientific: Cookbook



Original
(Brazilian
Portuguese)



**Translators'
first version**
(English)



**Revised
text**
(by native
speaker)



**Published
translation**
(not yet avail.
online)



Literary:

Australian short stories

(*learner corpus)



Original
(Australian
English)



**Student's
translation**
(Brazilian
Portuguese)



**Revised
draft** (after
teacher's
suggestions)



**Published
translation**



Search and annotation system

DISPARA (Santos 2002) – system to make parallel corpora available on the Web

✓ **Corpus processing system**

→ IMS-CWB (Christ *et al.* 1999), now **Open CWB** (Evert 2010)

✓ **Underlying parser and tagger**

→ Portuguese: **PALAVRAS** (Bick, 2000)

<http://visl.hum.sdu.dk/visl/pt/>

→ English: **CLAWS** (Rayson & Garside 1998)

<http://www.comp.lancs.ac.uk/computing/research/ucrel/claws/>

→ Semantic annotation: **corte-e-costura** (Santos & Mota 2010)

✓ **Interface** (graphic design by Patricia Tagnin)

CorTrad's Semantic annotation:

some findings on colour and clothing

Semantic information: colour

	Cooking	Scientific news	Short stories	Totals
Pure colour	574	372	344	1290
Conventional	310	153	2	465
Race	0	45	13	58
Human	0	7	39	46
Absence	7	21	22	50
Wine	87	1	6	94
Totals	985	599	428	
Word count	134,093	776,284	121,253	

Search expression: [sema="cor.*"]

Result type: semantic field

10 most recurring colour terms

Scientific news (or)		Short stories (tr)		Cookbook (or)	
cor	86	branco	69	dourar	341
branco	57	azul	48	branco	144
vermelho	43	vermelho	37	vermelho	122
verde	42	cor	26	verde	113
negro	37	negro	25	preto	39
<i>buraco negro</i>	31	preto	22	cor	39
<i>ultravioleta</i>	28	verde	19	tinto	33
<i>infravermelho</i>	27	amarelo	18	dourado	27
azul	25	pálido	15	amarelo	16
<i>amarelinho</i>	22	cinza	13	roxo	14

Search expression: [sema="cor.*"]

Result type: lemma distribution

“white” collocates in ≠ genres

Science news 45 (31)		Short stories 54 (27)		Cookbook 126 (24)	
glóbulo	4	mão	9	vinho	45
cabelo	4	homem	8	parte	14
cubo	3	pena	6	chocolate	12
mancha	3	cabelo	5	arroz	8
pelagem	2	látex	2	pele	8
luz	2	galão	2	fão de fôrma	7
população	2	blusa	2	milho	5
célula	2	...	1	pimenta-do-reino	3
...	1			fumaça	3
	1			carne	3
				peixe	3

Search expr.: [pos="N.*"] ([lema="branco"]|[grupo="Branco"]) Lemma distrib.

Simple search: “dour.*” is not “golden.*”



Confira as novas

Principal

O Projeto

Equipe

CorTec

CoMAprend

CorTrad

Artigos, etc.

Links

Informativo

Contato

Site FFLCH

Site USP

Início

Jornalístico

Literário

Técnico-científico

CorTrad técnico-científico culinária

O CorTrad é um corpus aberto, sujeito a alterações. Veja [dados quantitativos](#) para informações atualizadas sobre o conteúdo do corpus.

A **parte culinária do CorTrad** conta atualmente com o conteúdo integral de um livro de 350 páginas em quatro versões: o texto original, escrito em português brasileiro, a tradução para o inglês, feita por duas tradutoras brasileiras, a versão revisada dessa tradução, feita por uma falante nativa do inglês, e a versão final, publicada – esta, ainda em preparação. [Clique aqui para mais informações](#). A disponibilização do CorTrad na rede é um [projeto conjunto](#) entre o [COMET](#), a [Linguateca](#) e o [NILC](#), usando o sistema [DISPARA](#).

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal "dour.*"	<input checked="" type="checkbox"/> ver
Primeira tradução	<input type="radio"/> principal	<input checked="" type="checkbox"/> ver
Tradução revisada	<input type="radio"/> principal !"gold.*"	<input checked="" type="checkbox"/> ver

Ignorar maiúsculas/minúsculas

Pesquisar

Resultado

- | | |
|--|--|
| <input checked="" type="radio"/> Concordância | <input type="radio"/> Distribuição das formas |
| <input type="radio"/> Distribuição dos lemas | <input type="radio"/> Distribuição da categoria gramatical (PoS) |
| <input type="radio"/> Distribuição do tempo verbal e/ou do caso pronominal | <input type="radio"/> Distribuição de pessoa e/ou número |
| <input type="radio"/> Distribuição do gênero morfológico | <input type="radio"/> Distribuição da função sintática |
| <input type="radio"/> Distribuição por parte de obra | <input type="radio"/> Distribuição por tipo de texto |
| <input type="radio"/> Distribuição por receita | <input type="radio"/> Distribuição por campo semântico |

Options

Result: “dour.*” ≠ “golden.*”



	Português	Inglês	Português
Principal	Enquanto isso, doure ligeiramente os pinoli numa frigideira grande e seca e reserve.	Meanwhile, lightly toast pine nuts in a dry large frying pan and set aside.	Meanwhile, <u>lightly toast</u> pine nuts in a large dry frying pan and set aside.
O Projeto	Leve ao forno por uns 25 minutos para aquecer e dourar (se quiser, monte na véspera e guarde na geladeira sem assar).	Bake for about 25 minutes to heat thoroughly and brown the top (if you want, you can assemble lasagna a day ahead and refrigerate, unbaked).	Bake for about 25 minutes to heat thoroughly and <u>brown</u> the top (if you want, you can assemble lasagna a day ahead and refrigerate, unbaked).
Equipe	Regue com um fio de azeite e asse por uns 15 minutos, até que a massa esteja bem dourada e crescida nas bordas.	Drizzle with olive oil and bake for about 15 minutes, until dough is dark brown and risen on the edges.	Drizzle with olive oil and bake for about 15 minutes, until dough is <u>dark brown</u> and risen on the edges.
CorTec	Aqueça um fio de azeite numa panela grande que possa ir ao forno, junte metade da carne e mantenha o fogo forte para dourar e não juntar água.	Heat a drizzle of olive oil in a large ovenproof casserole dish, add half the meat and sear, over a high heat to prevent juices from escaping, until cubes are browned.	Heat a drizzle of olive oil in a large pot that can go into the oven, add half the meat and sear, over high heat to prevent juices from escaping, until cubes are <u>browned</u> .
CoMAprend	Transfira os pedaços dourados para um prato, aqueça mais um pouco de azeite, doure a carne restante e passe também para o prato.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.	Transfer <u>browned</u> chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.
CorTrad	Transfira os pedaços dourados para um prato, aqueça mais um pouco de azeite, doure a carne restante e passe também para o prato.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.	Transfer browned chunks to a plate, heat another drizzle of olive oil, sear remaining meat and transfer to the same plate.
Artigos, etc.	Aqueça mais um fio de azeite, junte o bacon e, quando começar a dourar , acrescente a cebola, a cenoura, o salsão e 1 pitada de sal.	Heat a little more olive oil, add bacon and, when it begins to brown, stir in diced onion, carrot, celery, and a pinch of salt.	Heat a little more olive oil, add bacon and, when it begins to <u>brown</u> , stir in diced onion, carrot, celery, and a pinch of salt.
Links	Pincele o rolo com a gema e asse por uns 25 minutos, até começar a dourar .	Brush roll with egg yolk and bake for 25 minutes or so, until it begins to brown.	Brush cylinder with egg yolk and bake for 25 minutes or so, until it begins to <u>brown</u> .
Informativo	Espalhe as fatias sobre a assadeira e volte ao forno por mais 20 minutos, virando na metade do tempo para dourar por igual.	Spread slices out on baking sheet and return to oven for another 20 minutes, turning once halfway through, so they brown evenly.	in) thick slices. Spread slices out on baking sheet and return to oven for another 20 minutes, turning once halfway through, so they <u>brown</u> evenly.
Contato	Para começar, como o clima é de muita descontração, vêm uns camarões de médios para grandes, passados no gergelim, dourados no forno e servidos com palitinhos com um molhinho agri-doce de abacaxi.	To start with, as the atmosphere exudes relaxation, there are the medium to large-sized shrimps, coated with sesame seeds, browned in the oven, and served with the aid of toothpicks for dipping in a sweet-n-sour pineapple sauce.	To start with, as the atmosphere exudes relaxation, there are the medium- to large-sized shrimps, coated with sesame seeds, <u>browned</u> in the oven, and served with the aid of toothpicks for dipping in a sweet-n-sour pineapple sauce.
Site FFLCH	Leve o salmão ao forno por uns 15 minutos, apenas o suficiente para terminar de dourar por fora e deixar o filé macio e rosado no centro.	Transfer salmon to preheated oven and bake for 15 minutes or so, just enough to crisp the outer crust and to tenderize the opaque pink flesh in the center.	Transfer salmon to preheated oven and bake for 15 minutes or so, just enough to <u>crisp</u> the outer crust and to tenderize the opaque pink flesh in the center.
Site USP	A segunda leva folhas verdes, manga e tirinhas douradas e picantes de peito de frango.	The second takes green salad leaves, mango and browned and spicy strips of chicken breasts.	The second takes salad greens, mango and <u>browned</u> and spicy strips of chicken breasts.
	Em seguida, aqueça um fio de óleo numa frigideira grande e	Then, discard chicken marinade, heat a drizzle of oil in a	Then, discard chicken marinade, heat a drizzle of oil in a large

Figurative expr. and terminology

- just declared itself **out to the blue**
- It was to be **black tie**.
- She never refused to go to Melbourne, but it was her hoodoo city, **a black jinx**.
- The dog knew they were coming, and barked **blue murder**.
- would quarrel with her till **the white hours**
- knowing about **brown rice**
- **blackfellow**
- **thin white sliced bread**
- **red wine**
- **red cabbage**

- surgira de repente, **do nada**
- O traje é **a rigor**.
- Nunca se negava a ir a Melbourne, mas era uma cidade de azar, **mau agouro**.
- O cachorro sabia que eles estavam vindo e latiu **desesperadamente**.
- de discutir com ela até **o amanhecer**
- Eu entendia sobre **arroz integral**
- **arborígene**
- **pão de forma**
- **vinho tinto**
- **repolho roxo**

Deemed unnecessary ... or different associations

- She rushed to the back of the house and hauled the drowsy **black** pup out of the kennel.
- She began to prowl between the desks, waving the **white** letter like a flag.
- When the torrent of **white** water subsided
- The **brown** smell of his cigar
- **white** and blue enamel bowls
- She had the **tinted** view of the Irish
- **golden** summer

- Ela correu para os fundos da casa e puxou o sonolento filhote para fora da casinha.
- Então começou a rondar as carteiras, balançando a carta como se fosse uma bandeira.
- Quando a corrente de águas espumantes cessou,
- O aroma característico do charuto
- tigelas azuis esmaltadas
- Ela tinha a visão **cor-de-rosa** dos irlandeses

Concluding remarks on colour

- Totally different translation patterns for
 - Figurative language (most cases do not preserve colour)
 - Skin/race/culture colour (more differentiation in English)
 - Real colour
- Scientific news: a lot of (unexpected) colour in scientific terminology: disease names, stars, etc.
- Short stories: high correlation of clothing and colour

10 most recurring clothing words

Scientific News (or)		Short Stories (tr)		Cookbook (or)	
Total oc.	58 (20)	Total oc.	380 (72)	Total oc.	21 (12)
roupa	17	roupa	37	roupa	7
sapato	9	chapéu	30	avental	2
luva	5	vestir	27	vestir	2
calçado	4	sapato	26	chapéu	2
camiseta	4	vestido	24	...	1
vestir	2	usar	22		
bota	2	camisa	13		
anel	2	meia	10		
vestido	2	calça	10		
...	1	capa	10		

Search expression: [sema="roupa.*"]

Result type: lemma distribution

Semantic information: clothing

	Cooking	Scientific news	Short stories	Totals
Pure clothing	21	50	378	449
Figurative	0	6	2	8
Totals	21	56	380	
Word count	134,093	776,284	121,253	

Search expression: [sema="roupa.*"]

Result type: semantic field

Figurative expressions and non-clothing-referring clothes

- The arrangement **suit**ed Henry extremely well.
- He says the government will set an example in **'belt-tightening'**.
- desafio de **se pôr nos sapatos** de um arqueólogo
- do tamanho de uma caixa de **sapatos**,
- pense na **roupa**

- O esquema **caiu como uma luva** para Henry.
- Ele diz que o governo estabelecerá um exemplo de **'apertar os cintos'**.
- task of **wearing the hat** of an archeologist
- the size of a **shoebox**
- think of what you will wear

Concluding remarks on clothing

- Also different translation patterns for
 - Figurative language
 - The issue of *wear* vs. *vestir/calçar* and *usar*
- Scientific news: expressions related to clothing but no clothes (Shoebox, washing machine)
- Short stories: how culture dependent are some descriptions, and what was the stance taken by the translators?

Corpus-based TS are revealing because

- you discover what were the challenges for the translator
- you discover vagueness in both languages
- you find problems you had never thought about before
- you find creative solutions

Acknowledgements

- Thanks to Eckhard Bick and Paul Rayson, for the use of PALAVRAS and CLAWS, respectively.
- Thanks to Sandra Aluísio and Arnaldo Candido Júnior at NILC for hosting and corresponding technical support.
- Thanks to Research Computing Services at Univ. Oslo
- This work was partially funded by the Portuguese government, UMIC, FCCN and the European Union (FEDER and FSE), under grant POSC/339/1.3/C/NAC (Linguatca)

