Explorando o CorTrad para pesquisa e(m) tradução

Diana Santos & Stella E.O. Tagnin

EBraLC - XI Escola Brasileira de Linguística de Corpus
d.s.m.santos@ilos.uio.no, seotagni@usp.br

23 de novembro de 2021

Plano da oficina

- Breve apresentação do CorTrad
- Questões lexicais e gramaticais
- Distribuições
- Buscas em campos semânticos específicos (cor, dizer, saúde, família)
- Buscas no original e na tradução simultaneamente

Breve apresentação do CorTrad, do ponto de vista da Linguateca

Idealizado como uma extensão do COMPARA

- em termos de géneros textuais
- em termos de múltiplas traduções
- em termos de várias versões de uma mesma tradução

Outros corpos relacionados, mas português-norueguês: PANTERA (como o COMPARA), e PoNTE (muitas traduções do mesmo texto).

Pressupostos teórico-metodológicos do CorTrad

- Não existe apenas uma tradução - a tradução é múltipla
- O processo de tradução passa por várias fases
- Um corpo linguístico permite encontrar resultados individuais e exemplos interessantes, mas também permite generalizar ao agregar casos – é o que significa **distribuição**

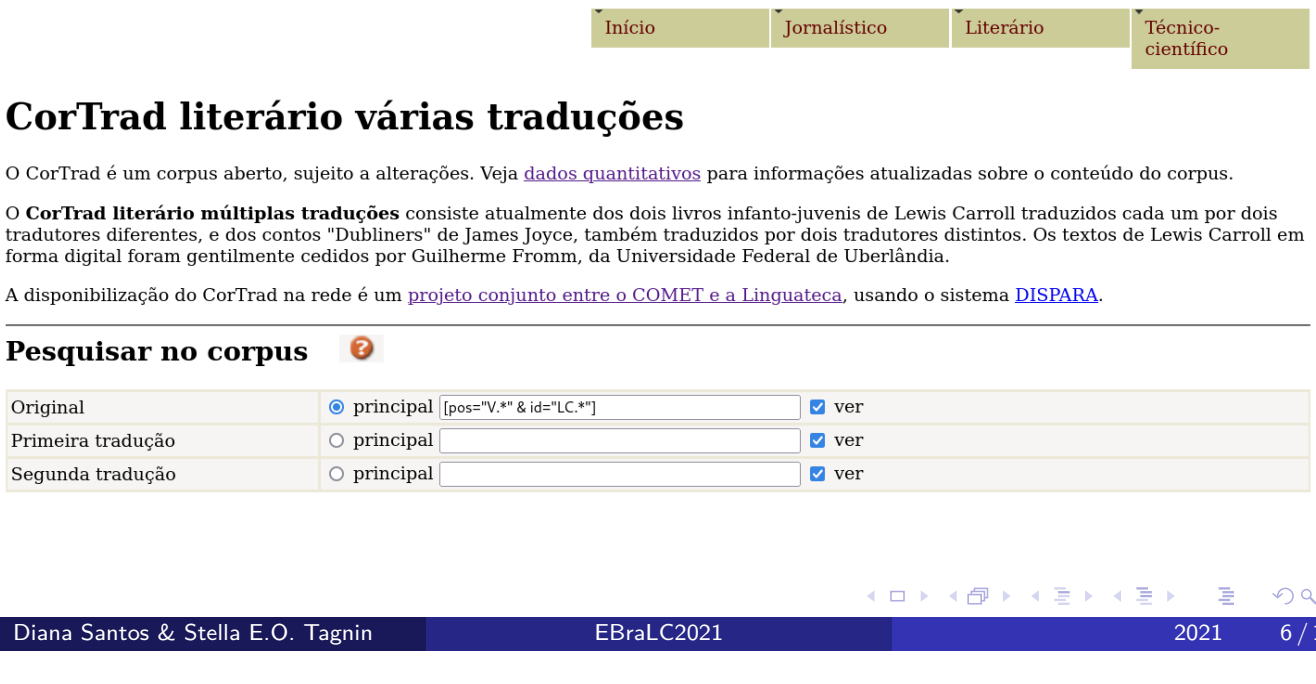
Informação preliminar sobre o formato de pesquisa

O CorTrad usa **expressões regulares**, para generalizar sobre as expressões de pesquisa.

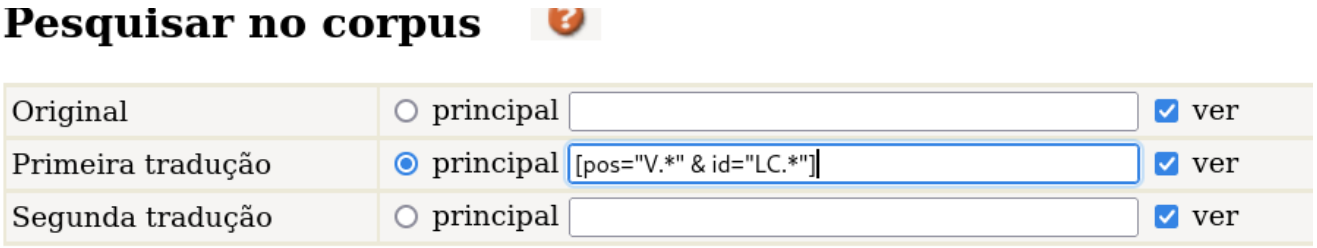
- Expressões regulares: Uma linguagem para descrever cadeias de caracteres (sequências de caracteres), com os seguintes caracteres especiais: `.`, `*`, `+`, `|`, e `[]` para definir classes de carateres.
- A sintaxe do OCWB: usa [ ] para as unidades, e elas próprias têm um conjunto de pares atributo-valor, como por exemplo para os atributos `lema` e `pos`...
- Pode-se usar expressões regulares dentro dos valores das unidades, ou sobre as unidades elas próprias.
- Quando há mais do que uma unidade, tem de se indicar onde começa cada uma. Com aspas, ou com [].

Quais os verbos mais frequentes que ocorrem na Alice?

- Primeiro, onde está a Alice?
- Depois, como a identificar?
- E como fazer a pergunta?



E nas traduções?



Campo semântico e grupo

- Campos semânticos (**sema**): áreas da experiência humana que têm o seu próprio vocabulário, e que muitas vezes também influenciam o resto da língua

cor	cor, cor:humana, cor:politica, cor:original, etc.
roupa	roupa
corpo	corpo, corpo:opiniao, corpo:animal, corpo:medida, etc.
relato	dizer, dizer-relatoDIRETO, dizer-relatoINDIRETO, etc.
emoções	emo:amor, emo:odio, emo:ausencia, etc.
parentesco	familia:lacos, familia:dinamica, familia:especificada
saúde	saude:doenca, saude:remedio, saude:medicina, etc.

- Grupos: para alguns campos semânticos, temos ainda mais uma "gaveta", mais uma coluna, chamada **grupo**. Para a cor, e para a roupa.

Veja-se `www.linguateca.pt/Gramateca/` (Lista de áreas)

Exemplos de cor

Procurar cor no CorTrad jornalístico.

- Que tipos de cor?
- Quando é cor mesmo, qual?

CorTrad jornalístico divulgação científica

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

A **parte jornalística do CorTrad** conta atualmente com textos da edições de 2001, 2002 e 2003 da Revista Pesquisa FAPESP, totalizando 20 números. As seções incluídas foram: Humanidades, Ciência, Tecnologia, Estratégias, Laboratório, Linha de Produção e Política de C & T. Veja uma tabela pormenorizada por assunto e gênero. A disponibilização do CorTrad na rede é um projeto conjunto entre o COMET e a Linguateca, usando o sistema DISPARA.

Pesquisar no corpus

Original	◉ principal	[sema="cor(:.*)*"]	☑ ver
Tradução publicada	○ principal		☑ ver

Exemplos de cor, continuação

- Em que tema é que há mais cores (cores mesmo)?

Original	◉ principal	[sema="cor"]	☑ ver
○ Distribuição por gênero de texto		◉ Distribuição por tema	

- Que palavras são mais modificadas por cores?

Original	◉ principal	[pos="N.*"] [sema="cor"]	☑ ver
◉ Distribuição dos lemas		○ Distribuição da categoria gramatical (PoS)	

- Quais os grupos de cor mais frequentes?

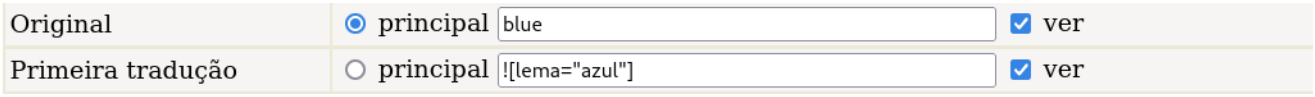
Original	◉ principal	[sema="cor"]	☑ ver
○ Distribuição por campo semântico		◉ Distribuição por grupo (de cor, de vestuário, etc.)	

Exercício: refazer a última procura no CorTrad culinário.

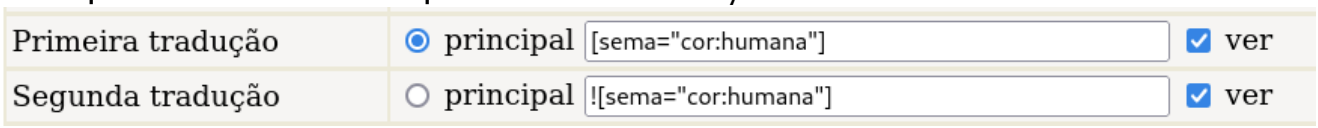
Traduções de *blue*

Usando agora o facto de que temos originais e traduções alinhados, podemos fazer procuras sobre a própria tradução:

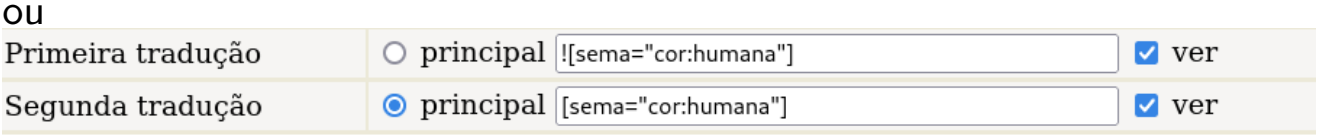
- Quais as traduções de *blue* nos Dubliners?
- Em que casos é que *blue* não é traduzido por *azul*?



- Em que casos de cor é que as duas traduções diferem?

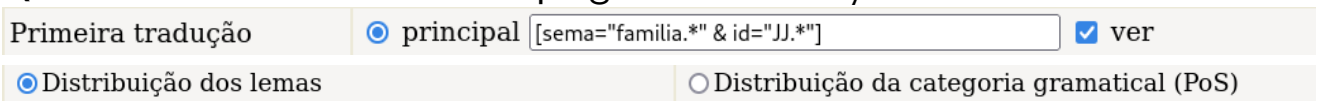


ou



A família nos *Dubliners*

- Quais os termos de família empregues nas traduções dos Dubliners?



Buscas relativas à saúde no corpo multiversão e no jornalístico

CorTrad jornalístico (

3360 casos.

Expressão de busca: **[sema="saude.*" %c]**
Resultado escolhido: **distribuição de lema**
Corpus pesquisado: **originais** (versão 4.0)

Houve 110 valores diferente

doença	500
bactéria	287
saúde	227
medicamento	224
médico	197
vírus	176
câncer	165
vacina	153
farmacêutico	89
infecção	81
infectar	68

CorTrad literário cor

157 casos.

Expressão de busca: **[sema="saude.*" %c]**
Resultado escolhido: **distribuição de lema**
Corpus pesquisado: **tradução final** (versão 6

Houve 44 valores diferentes

médico	21
doente	13
doença	11
hospital	11
depressão	10
saudável	8
ferimento	6
remédio	6
tosse	6
tuberculose	5
câncer	4

Buscas relativas a emoção

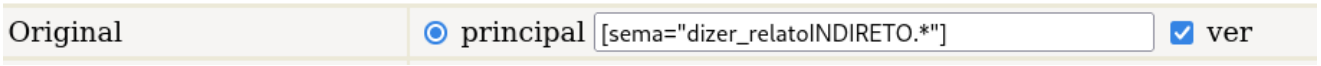
Identificar quando a revisão e a tradução publicada alteram o número de palavras de emoção... nos contos canadianos, há 1606 casos de emoção na primeira tradução, depois 1602 na tradução revisada, e 1604 na publicada...

Tradução revisada	○ principal	![sema="emo:.*"]	☑ ver
Tradução publicada	◉ principal	[colecao="CAN" & sema="emo:.*"]	☑ ver

Mrs Poorwilly waited until Sick Jack was covered, then brought Adel over to see Danny.	A sra.. Poorwilly esperou até Jack Doente ser coberto, e aí apresentou Adel a Danny.	Depois que Sick Jack foi agasalhado, a sra.. Poorwilly apresentou Adel a Danny.	A Sra. Poorwilly **esperou** que Sick Jack estivesse agasalhado e então apresentou Adel a Danny.
» When Danny didn't reply, Mrs Poorwilly added, «We got help moving your things.	Como Danny não respondesse, a sra.. Poorwilly acrescentou: Nós tivemos ajuda para retirar suas coisas.	Como Danny não respondesse, a sra.. Poorwilly acrescentou: Nós tivemos ajuda para fazer a sua mudança.	Como Danny não respondesse, a Sra. Poorwilly acrescentou: -- **Pedimos** ajuda para fazer a sua mudança.
The children solemnly nodded their heads.	As crianças assentiram de maneira respeitosa.	As crianças assentiram com a cabeça, de maneira solene.	As crianças assentiram **respeitosamente**.

Buscas relativas aos verbos dicendi

- Verbos que se referem a exprimir – campo lexical riquíssimo em português, estão marcados como `dizer`
- Verbos que exprimem um relato, que pode ser direto, indireto, ou misto, estão marcados como `dizer relato` e a seguir o tipo de relato



Buscas relativas ao tipo de alinhamento

- Atenção: o corpo jornalístico não tem essa informação!!!
- Frases que são traduzidas por várias frases

Não é preciso ter um motivo especial para convidar, basta querer matar a saudade de alguém.	You don't have to have a special reason to invite someone. Wishing to see someone you miss is as good a reason as any.	You don't have to have a special reason to invite someone. Wishing to see someone you miss is as good a reason as any.
Patricia Wells conta que, quando foi com o marido para a casa da Provence, achou que ficariam muito sozinhos.	Patricia Wells tells us of the time when she moved to the Provence with her husband. She thought they would feel very lonely.	Patricia Wells tells us of the time when she moved to Provence with her husband. She thought they would feel very lonely.

- Frases que são juntas na tradução em frases maiores

« What do you call yourself?» the Fawn said at last.	-- Como se chama você? perguntou ele por fim numa linda voz de veado.	-- Como se chama? -- perguntou por fim o animalzinho com a voz mais suave que Alice já ouvira.
Such a soft sweet voice it had!	-- Como se chama você? perguntou ele por fim numa linda voz de veado.	-- Como se chama? -- perguntou por fim o animalzinho com a voz mais suave que Alice já ouvira.

Comentários finais

Importante reter:

- Há muita coisa que não está perfeita: a maior parte da anotação semântica precisa de revisão!
- Nem todos devem concordar com todas as opções tomadas na anotação. A anotação implica interpretação por parte do anotador, não é algo "objetivo".
- Muitas vezes é possível modificar as pesquisas de forma a obter o que queremos, mas não "cai do céu"...
- O retorno de quem utiliza o CorTrad é **fundamental**!
- Há muita tarefas a distribuir por voluntários, se alguém quiser.

Mais informação

- Artigos sobre o CorTrad: `https://www.linguateca.pt/dispara/CorTrad/publicacoesCorTrad.php`
- Página de ajuda do CorTrad: `https://www.linguateca.pt/dispara/CorTrad/AjudaCorTrad.php`
- Informação sobre campos semânticos no AC/DC (e no CorTrad): `https://www.linguateca.pt/Gramateca/ListaAreas.html`
- Para outros corpos paralelos da Linguateca, procure no catálogo de publicações da Linguateca, com os marcadores COMPARA, PONTE, PANTERA, etc. : `https://www.linguateca.pt/superb/index.pl?load=busca_publ`