


Culinária brasileira: um CorTrad mais saboroso com o CulBras

Brazilian cooking: a tastier CorTrad with CulBras

Rozane R. Rebechi 
Instituto de Letras/PPG/UFRGS

Diana Santos 
Linguatca & ILOS, UiO

Stella E. O. Tagnin 
DLM/FFLCH/USP

Resumo

Embora muitos acreditem que traduzir receita culinária é algo corriqueiro, o tradutor experimenta problemas de diversos tipos ao se lançar a essa tarefa, desde a busca por equivalências lexicais até características culturais. Uma fonte de consulta que pode ser de grande auxílio é um corpus paralelo, ou seja, textos numa língua com suas respectivas traduções. O CorTrad culinário, lançado em 2008, numa parceria entre o COMET e a Linguatca, tem desempenhado esse papel, permitindo vários tipos de pesquisas. É composto das receitas de um livro escrito originalmente em português, acompanhado de suas traduções em inglês. Neste artigo apresentamos um novo corpus culinário, o CulBras, composto apenas por receitas brasileiras, originalmente escritas em português e suas respectivas traduções em inglês, totalizando aproximadamente 120.000 palavras em cada língua. O processo de construção desse corpus é apresentado em detalhes, assim como são relatadas várias pesquisas possíveis com os dois corpora, inclusive com a comparação entre eles. Convidamos os leitores a se deliciarem com o CulBras.

Palavras chave

linguística com corpora; corpora paralelos; tradução; culinária; culinária brasileira

Abstract

Although many believe that translating a culinary recipe is a straightforward task, translators face various challenges when undertaking this endeavor—from searching for lexical equivalents to navigating cultural nuances. A particularly useful reference source is a parallel corpus, that is, texts in one language alongside their respective translations. The culinary CorTrad, launched in 2008 as a partnership between Comet and Linguatca, has fulfilled this role, enabling various types of research. It consists of recipes from a book originally written in Portuguese, accompanied by their English translations. In this article, we present a new culinary corpus, CulBras, made up exclusively of Brazilian recipes originally written in Portuguese and their respective translations into English, totaling around 120,000 words in each language. The process of building this corpus is described in detail, as are several possible research avenues using

both corpora, including comparisons between them. We invite readers to savor CulBras.

Keywords

corpus linguistics; parallel corpora; translation; cooking; Brazilian cooking

1. A Importância da Tradução Culinária: Perspectivas Linguísticas e Culturais¹

A culinária é uma das expressões mais importantes da identidade cultural de um povo e a tradução de receitas permite que essa identidade seja compartilhada globalmente, fomentando o intercâmbio cultural e a possível adaptação de pratos tradicionais para diferentes públicos.

A despeito de seu papel como ponte cultural, a tradução de receitas não é trivial. Ela envolve desafios linguísticos, técnicos e culturais, como a adaptação de ingredientes e técnicas para diferentes contextos culturais e linguísticos. Vários desses aspectos vêm merecendo a atenção de acadêmicos (Chiaro & Rossato, 2015; Gaspari, 2015; Ghafarian et al., 2016).

Através da tradução, receitas tradicionais podem alcançar públicos internacionais, promovendo a valorização de ingredientes regionais e técnicas culinárias específicas. Isso contribui para a preservação de tradições e para a inovação gastronômica, à medida que diferentes culturas incorporam elementos de outras culinárias ao seu repertório. Um exemplo desse fenômeno é a disseminação do termo *caipirinha*, um aperitivo tipicamente brasileiro que hoje já é encontrado em várias culturas. O corpus de culinária brasileira (CulBras) ora apresentado comprova a manutenção do termo na tradução. O mesmo ocorre com o termo *cachaça*, que, embora difundido mundialmente como base da *caipirinha*, não tem sua tradução padroni-

¹Diferentes seções estarão em português do Brasil ou em português de Portugal, refletindo a diferente origem das autoras.



zada, conforme pesquisa num corpus bilíngue literário, o COMPARA (Frankenberg-Garcia & Santos, 2002). De dez ocorrências, o termo original é mantido apenas uma vez, sendo traduzido por *rum*, *crude brandy*, *sugarcane brandy*, dentre outras (Rebechi, 2015), corroborando a tendência domesticadora (cf. Venuti, 2008) das traduções literárias das línguas não hegemônicas para o inglês, à época em que foram publicadas. Vale ressaltar que as traduções no COMPARA datam de 1980 a 1990, enquanto os livros de receitas que compõem o CulBras foram publicados e traduzidos a partir de 2003.

Um dos problemas da tradução culinária são as medidas, pois os sistemas podem variar de um país para o outro. No Brasil prevalecem as medidas “a granel”, como *xícaras*, *colher de sopa*, enquanto em outros usam-se medidas precisas como *dl*, *ounces* etc. (Tagnin & Teixeira, 2004). Isso precisa ser levado em conta numa tradução.

Os ingredientes podem apresentar dois problemas: variação dialetal e disponibilidade regional. Tomemos, por exemplo, o termo *coentro* no português brasileiro que, no português de Portugal, ocorre no plural *coentros* (Tagnin & Teixeira, 2004). *Espinafre* é outro exemplo: singular no Brasil e plural em Portugal. O segundo problema são ingredientes típicos de uma cultura que não existem ou dificilmente são encontrados na outra, como é o caso de *buttermilk* no inglês. Embora exista uma tradução dicionarizada, *leitelho*, o produto não existe no Brasil, de modo que uma receita com esse ingrediente exigiria uma adaptação (Tagnin & Teixeira, 2004). Há ainda alguns acompanhamentos clássicos numa cultura que não existem na outra, como é o caso da *farofa* brasileira, das *migas* portuguesas ou dos *baked beans* norte-americanos (Tagnin & Teixeira, 2004).

Como se vê, com a globalização e o aumento da comunicação intercultural, a tradução de receitas culinárias se tornou uma ferramenta essencial para a promoção da diversidade cultural e gastronômica. Acreditamos que o corpus CulBras, ora apresentado, pode contribuir para uma comunicação culinária mais eficiente entre culturas diversas.

Este artigo tem como objetivo principal apresentar o corpus paralelo CulBras, composto por receitas originalmente escritas em português e suas traduções para o inglês, bem como discutir o seu potencial de aplicação em estudos de tradução e linguística de corpus. Para tanto, identificamos os termos e fraseologias mais recorrentes na culinária brasileira, bem como seus equivalentes em inglês, e examinamos as es-

tratégias tradutórias empregadas na tradução de nomes de receitas e expressões idiomáticas associadas à gastronomia brasileira e comparamos o léxico e o estilo de diferentes subcorpora culinários no CorTrad.

2. Os usos de corpora paralelos

Um corpus paralelo é aqui entendido como um conjunto de textos alinhados com uma ou mais traduções. O primeiro corpus paralelo usado em processamento de linguagem natural foram as transcrições do parlamento canadiano, os Hansards (Isabelle et al., 1993), nos anos 90 do século passado, usados para treinar um sistema de tradução automática (Brown et al., 1990). Mas cedo os corpora paralelos começaram a ser usados em muitos outros ramos do saber, em particular: estudos de tradução, estudos culturais, estudos contrastivos, estudos de ensino de línguas, e estudos linguísticos em geral. Alguns exemplos da variedade de estudos podem ser apreciados nos livros *Parallel Text Processing* (Véronis, 2000) e *Corpora na Tradução* (Viana & Tagnin, 2015).

O desenho dos corpora paralelos também se foi tornando mais sofisticado, tendo sido incorporados em sistemas maiores, que incluem corpora monolíngues, como o English-Norwegian Parallel Corpus, ENPC (Johansson, 2007), ou que apresentam múltiplas traduções ou diferentes versões de uma mesma tradução. Além disso, pode haver diversificação nas línguas (duas ou mais), assim como os corpora podem ser bidirecionais ou não, e conterem ou não traduções automáticas.

Seja como for, o desenho e a compilação destes corpora são geralmente mais complexos do que de corpora monolíngues, visto que exigem a identificação de pares (ou grupos) e, muitas vezes, o pedido de autorização a múltiplas entidades.

Visto que nem todos os tipos de textos são traduzidos, é também patente a maior escassez deste tipo de corpora, contrastando, por exemplo, com a existência de corpora comparáveis (ou seja, corpora de texto original com o mesmo tipo em várias línguas). Assim, o tipo de texto mais frequente em corpora paralelos são textos literários, legendas de filmes e transcrições de parlamentos ou outras instituições multilíngues, assim como resumos de textos acadêmicos, e a Bíblia. Exemplos de cada são RuN (Grønn & Marijanovic, 2010), OPUS (Tiedemann, 2009), Europarl (Kohn, 2002), SciPar (Roussis et al., 2022) e Bible corpus (Mayer & Cysouw, 2014). Contudo, também existem casos de tradução de notícias ou de textos não literários, como no CorTrad (ver próxima secção) ou no PANTERA (Santos,

2019). E, para o estudo específico do ensino de línguas e da tradução, existem corpora paralelos de aprendizes, como o PoNTE (Santos, 2014) e o trabalho descrito por Oliveira (2019).

Visto que cada um destes recursos descreve e contextualiza o tipo de investigação para que podem ser usados, deixamos a prova para a secção 5, onde demonstramos a utilidade do corpus aqui apresentado à comunidade.

Para um corpus paralelo recente relacionado com culinária, veja-se o trabalho de Haider et al. (2023), contendo legendas de programas culinários traduzidos do árabe para o inglês e vice-versa.

3. O CorTrad

O CorTrad teve início em 2008 graças a uma parceria entre o Projeto CoMET² e a Linguateca³, que implementou o DISPARA (Santos, 2002), um sistema para tornar corpora paralelos acessíveis online. O CorTrad consiste em textos originais em inglês e português alinhados com suas respectivas traduções. É composto de três gêneros: literário, técnico-científico e ciência popular. A direção da tradução dos textos literários é do inglês para o português, enquanto a dos outros dois gêneros é do português para o inglês.

Embora a ideia de um corpus paralelo remonte à concepção do Projeto CoMET (Tagnin, 2003), apenas em 2008 a cooperação com a Linguateca tornou o projeto viável. Foi anotado sintaticamente para o português utilizando o PALAVRAS (Bick, 2000), e etiquetado morfossintaticamente para o inglês primeiro pelo CLAWS (Rayson & Garside, 1998) e depois pelo TreeTagger (Schmid, 1994). Ao longo do tempo, o CorTrad passou por novos desenvolvimentos, incluindo anotações semânticas para cores, roupas e corpo (Santos et al., 2011, 2012) e, mais recentemente, para emoções, saúde, relações familiares e etnicidade.

O CorTrad consiste de três subcorpora, assim constituídos:

- Jornalístico (Português → Inglês): 1.076 textos da Revista Pesquisa FAPESP⁴, cujo objetivo é a divulgação da pesquisa científica feita no estado de São Paulo com o apoio desse órgão de fomento.
- Técnico-Científico (Português → Inglês): Livro de receitas (130.000 palavras), contendo

original, primeira tradução e tradução revisada, e agora enriquecido com o CulBras, tema do presente artigo.

- Literário (Inglês → Português): Subdividido em (a) Traduções e Revisões de contos australianos e canadenses, e (b) Múltiplas Traduções de obras de Lewis Carroll, James Joyce e Dr. Seuss.

A sintaxe das consultas pode ser bastante elaborada, dependendo do que está sendo investigado. Para facilitar o trabalho do pesquisador, alguns exemplos são mostrados na página inicial, referente ao subcorpus literário (Figura 1). Um botão de ajuda (um pequeno círculo vermelho com um ponto de interrogação) ao lado de “Pesquisar no corpus” leva o pesquisador mais aventureiro a uma lista maior de possibilidades de busca.

A Figura 1 mostra possíveis consultas para:

1. Palavra exata, como *down* neste caso
2. Palavras que começam com *viol*, exibindo ocorrências como *violin*, *violent*, *violet*, *violently* etc.
3. A expressão *at least*
4. Palavras que seguem dois pontos
5. Diferentes formas do verbo *see* em contexto
6. Palavras modificadas por *rather*
7. Casos em que a tradução não contém a palavra *vermelho*

Essa última funcionalidade é bastante útil para estudantes de tradução e pode gerar resultados interessantes, como o exemplo da Figura 2.

Isso demonstra que *red hair* não é traduzido como *cabelo vermelho*, mas sim como *cabelo ruivo*, uma colocação especializada em português. Contudo, note-se que na primeira tradução, antes da revisão, o tradutor foi levado a erroneamente escolher *cabelos vermelhos*.

Além das funções 4 e 6, que mostram os resultados pela distribuição de formas, os demais aparecem em linhas de concordância.

De fato, um dos aspectos mais notáveis do CorTrad é a possibilidade de apresentar múltiplas versões, pois a maioria dos textos apresenta mais de uma tradução, conforme ilustrado na Tabela 1, que mostra uma tradução inicial e uma versão revisada (as mudanças estão sublinhadas para facilitar a leitura).

A tabela evidencia mudanças feitas na versão revisada, como a reestruturação de *until a smooth paste is obtained* para *until it is a smooth paste*

²<http://comet.fflch.usp.br/>

³<https://www.linguateca.pt/>

⁴Fundação de Amparo à Pesquisa do Estado de São Paulo

	Procurar:	Resultado:
a palavra <i>down</i>	down	concordância
palavras começando por <i>viol</i>	"viol.*"	concordância
a expressão <i>at least</i>	"at" "least"	concordância
palavras antecedidas por dois pontos	": " @[]	distribuição das formas
formas do verbo <i>see</i> em contexto	[lema="see"]	concordância
palavras modificadas por <i>rather</i>	"rather" @[]	distribuição de formas
casos em que a tradução não tem <i>vermelho</i>	! [lema="vermelho"]	concordância

Figura 1: Captura de tela de algumas consultas sintáticas possíveis no CorTrad literário

He was about fifty with a round domed head, red hair beginning to go slightly white at the neck and ears, pink porcine skin, very neatly barbered and combed.	Ele beirava os cinquenta, tinha a cabeça redonda como um domo, os cabelos vermelhos começando a ficar levemente brancos no pescoço e orelhas, a pele um tom rosa porcino e estava barbeado e penteado com capricho.	Beirava os cinquenta, tinha a cabeça redonda como uma abóboda, os cabelos ruivos começando a ficar levemente brancos no pescoço e orelhas, a pele um tom rosa porcino e estava barbeado e penteado com capricho.	Beirava os cinquenta, tinha a cabeça redonda como uma abóbada, os cabelos ruivos começavam a ficar levemente brancos no pescoço e orelhas, a pele um tom rosa porcino e estava barbeado e penteado com capricho.
--	---	--	--

Figura 2: Procura de *red* no original e não *vermelho* na tradução final

e de *until a smooth and fluffy cream is obtained* para simplesmente *until smooth and fluffy*. Esse formato permite o estudo das diferentes etapas de uma tradução e do processo de revisão (Teixeira & Tagnin, 2008). Também auxilia os alunos a compreenderem soluções mais naturais na tradução.

Conforme mencionado anteriormente, no domínio da Culinária o CorTrad já contava com o livro de receitas *Cozinhando para Amigos* (Bacellar, 2008), doravante CpA, com cerca de 130.000 palavras, escrito originalmente em português, traduzido por duas tradutoras e finalmente revisado por uma falante nativa americana.

Esse corpus permite buscas por partes específicas do documento como, por exemplo, ingredientes, preparo, comentários etc., ou ainda por receita.

4. O CulBras

4.1. Conteúdo

A seleção do corpus paralelo de receitas culinárias brasileiras (CulBras) nesta pesquisa fundamentou-se na proposta de identificar os termos e fraseologias mais recorrentes em receitas culinárias brasileiras e encontrar seus equivalentes em inglês, com vistas à construção de uma obra de referência para tradutores e redatores especializados na área da gastronomia, especialmente a brasileira. Esse corpus foi construído no âmbito da tese que teve como objetivo a construção de um glossário de termos da culinária brasileira em português e inglês (Rebechi, 2015). Os critérios de seleção incluíram livros publica-

dos entre 2003 e 2014, com o objetivo de evitar o uso de vocabulário obsoleto, e cujos títulos remetessem à culinária brasileira de maneira ampla (referências ao Brasil como um todo, a regiões ou a estados específicos). Os livros selecionados foram originalmente escritos em português e traduzidos para o inglês, independentemente de os textos em ambos os idiomas estarem contidos em um único volume ou em volumes distintos.

A maior parte do material estava disponível apenas em formato impresso, com exceção de um dos volumes, disponível em formato digital. Por isso, foi necessário digitalizar o material impresso para que todas as receitas pudessem ser processadas por ferramentas computacionais. A digitalização foi realizada com o uso de scanner de alta resolução e softwares de OCR (reconhecimento óptico de caracteres). Esse processo revelou-se especialmente trabalhoso devido ao design gráfico dos livros de receitas, geralmente coloridos e com grande diversidade tipográfica, o que causava erros frequentes no reconhecimento de caracteres – por exemplo, a confusão entre “m” e “nr”, resultando em palavras distorcidas – por exemplo, *came* em inglês sendo lido como *carne* ou *com* em português como *corn*. Assim, a etapa posterior de revisão manual foi fundamental para garantir a integridade dos dados, pois as distorções textuais poderiam comprometer os resultados das análises automáticas.

Embora os livros utilizados conttenham não apenas receitas, mas também glossários, textos introdutórios, curiosidades e informações culturais, o corpus paralelo foi estruturado com foco

Original	Primeira tradução	Tradução revisada
Deixe amornar e bata no liquidificador com os ovos e o creme de leite até obter uma pasta lisa, acerte o sal e a pimenta, junte a salsinha, a cebolinha e o tomilho e reserve.	Let cool until warm and then combine eggplant, eggs and whipping cream in a blender <u>until a smooth paste is obtained</u> ; adjust salt and pepper, fold in chopped parsley, chives and thyme leaves and set aside.	Let cool until warm and then puré eggplant, eggs and whipping cream in a blender until it is a smooth paste; adjust salt and pepper, fold in chopped parsley, chives and thyme leaves and set aside.
Prepare o creme com pelo menos 2 horas de antecedência: com a batedeira, bata o creme de leite, o cream cheese, a baunilha e o açúcar mascavo até obter uma pasta lisa e macia, cubra e guarde por até 2 dias na geladeira.	Whip the cream at least 2 hours before serving: in an electric mixer, combine whipping cream, cream cheese, vanilla extract, and brown sugar, beat <u>until a smooth and fluffy cream is obtained</u> , cover with plastic wrap (cling film) and refrigerate for up to 2 days.	Whip the cream at least 2 hours before serving: in an electric mixer, combine whipping cream, cream cheese, vanilla extract, and brown sugar, beat <u>until smooth and fluffy</u> , cover with plastic wrap (cling film) and refrigerate for up to 2 days.

Tabela 1: Traduções de *até obter uma pasta...*

exclusivo nas receitas.⁵ O motivo principal para tal escolha reside no objetivo de buscar equivalentes lexicais, terminológicos e fraseológicos diretamente ligados ao gênero receita culinária.

As receitas foram agrupadas por livro e salvas separadamente. Por exemplo, BGCTarroz03-P.txt se refere à terceira receita de arroz encontrada no livro *Brasil: Gastronomia, Cultura e Turismo*, salva em formato de texto sem formatação. Sua contraparte em inglês foi salva como BGCTarroz03-I.txt.

O corpus paralelo resultante da compilação totalizou 161.027 palavras em português e 165.665 em inglês. Vale ressaltar que estes números não coincidem exatamente com o conteúdo final do CulBras no CorTrad porque excluímos algumas receitas que não tinham sido traduzidas. Além disso, as ferramentas de contagem de palavras são diferentes (veja-se Grefenstette & Tapainen (1994); Santos et al. (2003)).

As seis obras incluídas no CulBras foram:

BGCT Eva Ribenboim Steinbruch. *Brasil: Gastronomia, Cultura e Turismo* BEI, São Paulo, 2010. Traduzido por: Rose Veelemon, Daniela Sandle e Anthony Doyle, *Brazil: Gastronomy, Culture and Tourism*.

CBCC Elisa Lopes Campos. *Culinária Brasileira, com Certeza*. Edição eletrônica, 2012. Traduzido por: Elisa Lopes Campos, *Brazilian culinary, certainly: everyday recipes*.

CdO Thiago Castanho e Luciana Bianchi. *Cozinha de Origem*. PubliFolha, São Paulo, 2014. Traduzido por: Elisa Duarte Teixeira, *Brazilian Food*.

SB André Boccato. *Sabores Brasileiros: 40 receitas típicas*. Editora Boccato, São Paulo, 2013. Traduzido por: Lauren A. B. Dobbin, *Brazilian Taste: 40 Typical Recipes*.

SSC Delfina Rocha. *Sabores e Saberes do Ceará*. DM Rocha, Fortaleza, 2003. Traduzido por: Graeme Clive Hodgson.

VGAB Caloca Fernandes. *Viagem Gastronômica Através do Brasil*. Editora Senac São Paulo, São Paulo, 2007. Traduzido por: Doris Hefti.

4.2. Incorporação no CorTrad

Cada receita correspondia a um ficheiro, marcado em pseudo-XML com as seguintes secções: <titrec> título da receita, <ingr> (contendo todos os ingredientes, um por linha), <modfaz> contendo o modo de fazer, <class> indicando o número de porções a que a receita se referia, assim como eventualmente o tempo de preparação, <coment> comentários que em geral se referem à origem da receita e/ou à sua disseminação geográfica, e <foto> para legendas de fotos.

Cada par de ficheiros foi então, depois de algumas reformulações automáticas, como a remoção dos cabeçalhos, e dos * que indicavam início de ingrediente nalgumas obras, alinhado automaticamente com o programa cwb-align do IMS Open

⁵ Algumas receitas eram acompanhadas de comentários sobre sua origem, autoria, entre outros aspectos. Essas informações foram preservadas.

Início
Jornalístico
Literário
Técnico-científico

CorTrad técnico-científico culinária brasileira

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

A **parte culinária brasileira do CorTrad** conta atualmente com o conteúdo integral de seis livros de cozinha brasileira, traduzidos para o inglês dos Estados Unidos. [Clique aqui para mais informações](#).

A disponibilização do CorTrad na rede é um [projeto conjunto entre o COMET e a Linguateca](#), usando o sistema [DISPARA](#).

Pesquisar no corpus ?

Original	<input checked="" type="radio"/> principal <input style="width: 100px;" type="text"/>	<input checked="" type="checkbox"/> ver
Tradução publicada	<input type="radio"/> principal <input style="width: 100px;" type="text"/>	<input checked="" type="checkbox"/> ver

☒ Ignorar maiúsculas/minúsculas

Procurar por tipo de alinhamento:

☐ 1-0
 ☐ 1-N
 ☐ N-1
 ☐ N-M

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 campo semântico	<input type="radio"/> Distribuição por grupo (de cor, de vestuário, etc.)
<input type="radio"/> Distribuição pelas molduras	<input type="radio"/> Distribuição pelas obras
<input type="radio"/> Distribuição por tipo de texto	

Opções

☐ Resultados por ordem alfabética (só distribuições)
☐ Resultados numerados (só concordância)

Pesquisar

Figura 3: A interface de procura no CulBras

CWB (Evert & Team, 2022), e concatenados em dois ficheiros, um para português e outro para inglês. O alinhamento foi depois cuidadosamente revisto, garantindo as seguintes características, comuns aos outros corpora paralelos disponibilizados pela Linguateca, e iniciadas no âmbito do COMPARA (Frankenberg-Garcia & Santos, 2002), nomeadamente:

- uma frase no lado original (neste caso o português)
- no lado traduzido, uma ou mais frases correspondentes, ou apenas uma parte da frase⁶
- no caso de frases adicionadas na tradução, estão marcadas como <adic>, adições, mas encontram-se na unidade de alinhamento anterior

⁶No COMPARA, quando a tradução era de N para 1, repetíamos a frase completa N vezes do lado traduzido, mas, noutros corpora paralelos, nomeadamente no PONTE (Santos, 2014), passámos a dividir a frase traduzida em N partes.

- no caso de frases sem tradução (1-0) colocamos uma frase vazia (na interface aparece depois a informação **ABSOLUTAMENTE NADA!!!**)

Além disso, alguns casos de receitas que não tinham sido traduzidas foram retirados, e pequenas gralhas ou problemas de digitalização foram sendo corrigidos.

Depois de revistos, estes dois ficheiros seguiram o processo do CorTrad: o lado português foi analisado sintaticamente pelo PALAVRAS (Bick, 2014), e anotado com os campos semânticos estudados na Linguateca: cor, roupa, corpo, saúde, emoções e comida/bebida. O lado inglês é anotado pelo TreeTagger (Schmid, 1994) e adicionada informação sobre a cor. Depois são codificados no sistema Open CWB (Evert & Team, 2022) e disponibilizados para interrogação na rede através do sistema DISPARA (Santos, 2002), no sítio do CorTrad (ver Figura 3).

13. <s> </ListaIngred> <ModoDeFazer> Es-corra e descasque o feijão e passe-o na máquina de moer carne (com o furo liso) . </s>	13. <s> </ListaIngred> <ModoDeFazer> Place the black-eyed peas in a bowl, cover with water, and soak overnight. The following day, remove the skins from the black-eyed peas and discard. Finely grind the peeled black-eyed peas in a meat grinder (using the fine grinding plate), until it forms a homogeneous paste.
423. <s> </ModoDeFazer> <Class> 80 unidades </s>	423. <s> </ModoDeFazer> <Class> Makes 80 cookies </Class> <adic> <Coment> Cuceca virada literally means "inside-out briefs ". </adic>
474. <s> </ListaIngred> <ModoDeFazer> Retire as castanhas dos cajus . </s>	474. <s>
2167. <s> 2- Junte o arroz já pronto e a cebolinha e misture bem . </s>	2167. <s> 2- Add the prepared rice and spring onions. Stir well to combine
2168. <s> Sirva imediatamente . </s>	2168. <s> and serve immediately.
2172. <s> <ingred> 150 g de aviú ou camarão seco bem picado </ingred> </s>	2172. <s> <ingred> 150g dried aviú (tiny dried, salty shrimps) or regular dried salt shrimps </ingred>

Tabela 2: Exemplos do formato, depois da revisão de alinhamento. Cada linha da tabela corresponde a uma unidade de alinhamento. As marcações entre < e > correspondem a marcações internas de estrutura, já existentes, e <s> e </s> indicam respetivamente o início e o fim de uma frase.

Na Tabela 3 mostra-se uma panorâmica do conteúdo do corpus, que evidencia diferentes estilos de cada obra, e também dos tradutores. Analisaremos estas diferenças em mais pormenor em secções posteriores.

Embora mais de metade das receitas tivesse algum tipo de comentário, apenas havia legendas de fotos traduzidas em VGAB (37 casos) e SSC (6 casos).

Quanto a identificar as receitas comuns (ou seja, as que aparecem em mais de um livro), e que podem assim indicar as receitas mais brasileiras (porque repetidas), os resultados estão na Tabela 4.

Mais interessante é confirmar que diferentes tradutores usaram diferentes estratégias para traduzir os nomes, e, portanto, no lado inglês o número de receitas traduzidas semelhantes (se só verificássemos pelo nome) é muito menor, como se pode apreciar na Tabela 5.

Na secção 5.2 discutiremos o interessante problema da tradução dos títulos.

5. Exemplos de pesquisas

5.1. Culinária brasileira vs. culinária em geral

Os dois corpora de culinária (o CorTrad culinária várias versões e o CulBras) são aproximadamente do mesmo tamanho (ver tabela 6, por isso podemos comparar diretamente as frequências absolutas, para identificar quais os termos mais frequentes na culinária brasileira vs. os termos mais empregues na culinária usada no Brasil, mas não necessariamente brasileira. Claro está que esta comparação pressupõe que o livro de culinária *Cozinhando para amigos* seja representativo da culinária que se faz no Brasil e não só da autora, o que não podemos afirmar nem naturalmente confirmar. De facto, segundo Elisa Duarte Teixeira, o CpA não é de forma alguma representativo da culinária no Brasil, é um livro altamente pessoal, organizado por refeições ou temas, que incluem várias receitas, e que tem muito palavreado incomum num texto de culinária. Ainda segundo esta especialista da culinária brasileira e sua tradução, um estudo mais sólido seria a comparação do CulBras com a parte culinária do CorTec. Contudo, mantivemos o exemplo no ar-

Obra	Receitas	UAs	U port	U ing	Palavras/receita port	Palavras/receita ing
BGCT	60	1213	11181	11928	186	198
CBCC	80	925	8403	8325	105	104
CdO	112	3262	40010	38164	357	340
SB	40	988	9488	7573	237	189
SSC	39	838	9996	9892	256	253
VGAB	135	3327	44203	43363	327	321
Total	466	10554	123305	119259	265	234

Tabela 3: Tabela panorâmica do conteúdo do CulBras, versão 2.4 (30 de junho de 2025): UA - unidades de alinhamento, U - unidades

Acarajé	3
Ambrosia	2
Arroz com guariroba	2
Arroz de carreteiro	4
Arroz de cuxá	2
Atolado de caranguejo	2
Barreado	3
Bobó de camarão	2
Bolinho de arroz	2
Bolo de Fubá	2
Bolo de macaxeira	2
Caldo de peixe	2
Canjica	2
Caruru	3
Casquinha de caranguejo	3
Cuscuz paulista	2
Doce de cupuaçu	2
Doce de limõezinhos	2
Empadão goiano	2
Farofa de banana	2
Feijoadá	2
Feijoadá completa	2
Furrundum	2
Galinhada	2
Maniçoba	2
Pato no tucupi	2
Pirarucu de casaca	2
Pão de Queijo	2
Tacacá	3
Torta capixaba	2
Vaca atolada	2
Vatapá	2

Tabela 4: Receitas repetidas no CulBras

Ambrosia	2
Baked sweet potatoes with clarified butter	2
Barreado	2
Cheese rolls	2
Cupuaçu paste	2

Tabela 5: Receitas repetidas no lado inglês

tigo porque demonstra o tipo de estudos que se podem fazer com o CorTrad. Aliás, é preciso dizer que, na tese em cujo âmbito o CulBras foi desenvolvido (Rebechi, 2015), houve estudos mais completos sobre exatamente essa questão, comparando a culinária brasileira a corpora de livros de culinária geral no Brasil, tanto em português como em inglês. Por isso, a comparação aqui descrita apenas complementa esse estudo maior já feito e publicado, e que usou outras ferramentas de corpora, como WordSmith (Scott, 2012).

De qualquer maneira, é interessante indicar quais as hipóteses que pretendíamos verificar antes de executar a comparação propriamente dita:

- mais consistência/menos diversidade lexical no livro *Cozinhando Para Amigos*
- menos palavras geográficas e de nacionalidade
- mais conversa com o leitor sobre o contexto em que um prato é cozinhado ou servido
- mais estrangeirismos (por causa da cozinha internacional)

A primeira expectativa foi verificada: há mais lemas diferentes na culinária brasileira do que no livro CpA, muito provavelmente porque autores diferentes usam palavras diferentes, e também provavelmente por causa das explicações botânicas e culturais, inexistentes no CpA.

Quanto às outras hipóteses, o mais fácil é obter (primeiro) os itens lexicais com maior diferença de frequência entre os dois corpora, tendo em conta lemas com mais de 10 ocorrências no total dos dois corpora. A Tabela 7 mostra os trinta maiores desvios, medidos como a diferença de frequências sobre a frequência total em ambos os corpora. Ou seja, se o verbo *aquecer* aparece 551 vezes no CpA e apenas 276 vezes no CulBras, a sua frequência total é $551+276=827$ e a sua diferença de frequências é $551-276=275$. Assim, o valor do desvio é 0,332. Se uma palavra aparecer mais no CulBras do que no CpA, o seu desvio é negativo.

Português				Inglês		
corpus	unidades	tipos	lemas	unidades	tipos	lemas
CpA	138649	7528	4428	159864	6935	4926
CulBras	123305	7463	4656	119259	6278	4829

Tabela 6: Dimensões dos dois corpora culinários no CorTrad

Lema	Dif	CpA	CulBras
cogumelo	1,000	44	0
macaxeira	-1,000	0	47
brando	-1,000	0	55
piracuru	-1,000	0	46
corrente	-1,000	0	33
goma	-1,000	0	42
iogurte	1,000	40	0
pinoli	1,000	30	0
pistache	1,000	48	0
rendimento	1,000	0	149
cheiro-verde	-1,000	0	88
damasco	1,000	43	0
tapioca	-1,000	0	81
berinjela	1,000	48	0
lindo	1,000	41	0
caranguejo	-1,000	0	65
tucupi	-1,000	0	62
endro	1,000	30	0
urucum	-1,000	0	32
interessante	1,000	39	0
posta	-1,000	0	30
pimenta-de-cheiro	-1,000	0	89
chuchu	-1,000	0	31
colorau	-1,000	0	58
charque	-1,000	0	42
jambo	-1,000	0	56
perfumar	0,9821	111	1
preaquecer	-0,9753	1	80
coisa	0,9722	71	1
amassar	-0,9690	2	127
abobrinha	0,9643	55	1
oliva	0,9627	158	3
montar	0,9600	49	1
deitar	-0,9600	1	49
freezer	0,9592	48	1
toucinho	-0,9574	1	46
ajustar	0,9555	88	2

Tabela 7: Diferenças lexicais entre os dois corpora

Estas primeiras diferenças lexicais não nos permitem avaliar as hipóteses que mencionamos, mas mostram algumas diferenças de estilo: *ajustar*, *montar* e *perfumar* (como verbos de culinária) são muito mais usados no CpA; já *preaquecer* (forno preaquecido), *amassar* (geralmente dentes de alho) e *deitar* (apenas no VGAB, referindo-se a um ingrediente líquido) são quase exclusivas do CulBras.

Quanto a ingredientes, por esta simples comparação pode ver-se que *macaxeira*, *piracuru*, *cheiro-verde*, *tapioca*, *caranguejo*, *tucupi*, *urucum*, *pimenta-de-cheiro*, *chuchu*, *colorau*, *charque*, *jambo* e *toucinho* devem ser típicos da culinária brasileira, enquanto *cogumelos*, *iogurte*, *pinoli*, *berinjela*, *endro* (sempre indicado, no CpA, com o nome inglês entre parênteses: *endro* (*dill*)) e *abobrinha* provavelmente não.

Outra descoberta mais inesperada é o uso de *fogo brando* e de *água corrente* no CulBras, mas nunca no CpA. Já a palavra *rendimento* é apenas a forma de indicar o número de porções que a receita produz, em dois dos livros do CulBras. *Oliva* aparece sempre em *azeite de oliva*, apenas no CpA. Finalmente, palavras como *interessante* e *coisa*, muito mais usadas no CpA, encontram-se exclusiva ou majoritariamente na parte de introdução e não nas receitas. E relembramos que o CulBras apenas inclui as receitas, e não palavreado sobre culinária.

Vejamos então como indagar se o CulBras contém mais palavras geográficas: procurando pela análise semântica do PALAVRAS, localizações. Na parte do recheio do CpA, obtivemos 103 casos, enquanto que no CulBras obtivemos 255. Embora muitas vezes a análise seja incorreta, parece confirmar as nossas expectativas, sobretudo se compararmos os nomes próprios:

No CulBras: *Belém*, *Minas Gerais*, *São Paulo*, *Rio de Janeiro*, *Mato Grosso*, *Goiás*, *Manaus*, *Pedra Azul*, *Santa Catarina*, *Pará*, *Sobral*, *Fortaleza*, *Porto Seguro*, *Rio Grande do Sul*, *Salvador*, *Santarém*, *Antonina*, *Bragança*, *Lagoa da Conceição*, *Maracaju*, *Espírito Santo*, *Goiás Velho*, *Goiânia*, *Pernambuco*, *Porto Feliz*, *São João*, *São Luís*, *Santa Terezinha*, *Tietê*, *Trancoso*, *Açores*, *Alagoas*, *Alcântara*, *Alegrete*, *Beiras*, *Ceará*, *Curitiba*, *Florianópolis*, *Goa*, *Goia-*

tuba, Ipanema, Jaguaribe, Lapa, Lisboa, Morretes, Morungaba, Paraná, Paranaguá, Piauí, Recife, Ribatejo, Serro, Sintra, Tocantins, Uberaba. Apenas três se referem a locais fora do Brasil, em Portugal e na Índia.

Pelo contrário, na obra CpA, temos *Dijon, Paris, Porto, Bordeaux, Guérande, Nova York, Rio Grande do Norte, San Francisco, Cairo, Avignon, Bretanha, Califórnia, Estados Unidos, João Pessoa, Minas Gerais, Normandia, Nova Inglaterra, São Paulo, Alentejo, Amsterdã, Andaluzia, Andes, Belém, Brasília, Carpentras, Castelnau-dary, Damasco, EUA, Gordes, Honfleur, Locronan, Petaluma, Roma, Sevilha, Toscana, San Gimignano, Volterra, Borgonha*. A grande maioria são locais fora do Brasil.

Finalmente, a questão dos estrangeirismos é difícil de quantificar sem observar todo o léxico, mas basta verificar que entre os títulos das 356 receitas do CpA há 120 que incluem pelo menos um termo estrangeiro, ao contrário dos da culinária brasileira.

5.2. A tradução dos nomes das receitas

Os títulos das receitas culinárias funcionam como “resumos” nesse gênero textual (Labov, 1973), já que servem para instigar e informar o leitor. No que tange aos nomes das receitas brasileiras que compõem o corpus paralelo, Tagnin et al. (2022) os classificaram em 5 categorias: informativos, com marcas geográficas, com epônimos, idiomáticos e avaliativos.

Por meio de uma busca no CulBras, identificamos os títulos de todas as receitas do corpus, alinhados às traduções. A Figura 4 apresenta uma amostra dos resultados.

Conforme discutido em Tagnin et al. (2022), muitos títulos de receitas da culinária brasileira possuem caráter idiomático, ou seja, não podem ser compreendidos pela simples soma de seus elementos lexicais. Expressões como *Baba de Moça*, *Pé-de-Moleque* ou *Baião-de-Dois*, se interpretadas literalmente, remetem a imagens que nada têm a ver com comida. No entanto, no contexto gastronômico, essas expressões são entendidas como unidades fixas de significado, consolidadas pelo uso (Tagnin, 2013). A tradução literal desses títulos pode causar estranhamento ao leitor estrangeiro, que tende a interpretá-los composicionalmente. Um exemplo representativo desse desafio é o caso do *arroz de carreteiro*. Descrito (como receita) cinco vezes no CulBras, é um prato típico da culinária gaúcha brasileira, tradicionalmente preparado com arroz e charque (ou carne seca). Seu nome remete diretamente

aos trabalhadores que conduziam carretas puxadas por bois e percorriam longas distâncias transportando mercadorias no sul do Brasil durante o período colonial e imperial. Por ser um prato de preparo relativamente simples e com ingredientes duráveis, era uma opção prática para esses viajantes — daí sua associação cultural e etimológica ao termo *carreteiro*.

Na tradução para o inglês, encontram-se diversas estratégias: o empréstimo direto (em *Carreteiro rice*), a tradução literal com falso cognato (em *Carter rice*), a descrição dos ingredientes (em *Rice with Brazilian jerk beef, pork ribs and sausage*), e as adaptações culturais (em *Cowboy's rice* ou *Mule-driver's rice*). Cada uma dessas opções evidencia diferentes abordagens tradutórias diante de um título consolidado idiomáticamente na língua de partida.

Observa-se também uma diversidade de estratégias tradutórias nos títulos que fazem referência a ingredientes tipicamente brasileiros, como *cuxá*; a preparos específicos, como *moqueca*; a variações regionais de um mesmo prato, como *cuscuz paulista*; ou ainda a nomes de personalidades, como *Filé à Oswaldo Aranha*.

5.3. Diferenças entre os seis livros

Na tabela panorâmica do conteúdo do CulBras, já tinha sido possível apreciar que as obras constantes do corpus são díspares em termos do seu tamanho, do número de receitas que incluem, do número de palavras por receita e da quantidade de comentários incluídos. Contudo, é possível fazer outro tipo de micro-estudos, dada a anotação fina a que as receitas foram submetidas. Por exemplo, é possível contar o número de ingredientes por receita, por obra, como ilustrado na Figura 6.

Podemos assim afirmar que a obra SB tem as receitas mais elaboradas (ou pelo menos com maior número de ingredientes em média), sendo a obra CBCC a que tem as receitas com menos ingredientes.

Também é possível identificar quantas receitas são doces ou salgadas, através da simples contagem do ingrediente sal e do ingrediente açúcar. (Outras procuras mais complicadas ficam para os leitores e/ou utilizadores do CulBras.)

Assim, vemos que existem 322 receitas em que sal está mencionado como ingrediente, contra apenas 155 que contêm açúcar (interessante, apenas uma receita inclui mel, os fartes quinhentistas na obra VGAB). Na Figura 5, mostramos as primeiras receitas com o ingrediente açúcar, três vezes. O que significa que esse ingrediente integra

Expressão de busca: <t> [* "arroz" %c [* </t>
 Resultado escolhido: **concordância em contexto**
 Corpus pesquisado: **originais** (versão 2.7)



30 ocorrências.

[Voltar](#)
[Nova pesquisa](#)

Original	Tradução publicada
Arroz de cuxá	Cuxá rice
Arroz de hauçá	Hauçá rice
Arroz com guariroba	Rice with guariroba
Arroz com pequi	Rice with pequi
Arroz de carreteiro	Carreteiro rice
Bolinho de arroz	Rice cupcakes
Arroz de forno	Rice in the oven
Arroz de carreteiro	Carter rice

Figura 4: Títulos de receitas incluindo arroz (primeiras sete ocorrências)

CorTrad técnico-científico culinária brasileira

190 casos.

Expressão de busca: <ingred> [* [lema="açúcar|mel" %c] [* </ingred>
 Resultado escolhido: **distribuição de receita**
 Corpus pesquisado: **originais** (versão 2.7)



Houve 155 valores diferentes de receita.

[Voltar](#)
[Nova pesquisa](#)

BGCT-Torta de banana-pacová	3
BGCT-Torta de nozes	3
CBCC-Ameixa de queijo mineiro	3
CdO-Bolo de chocolate com cupuaçu	3
CdO-Tiramisù de bacuri	3
SB-Torta de Castanha do Pará com Cupuaçu	3
VGAB-Musse de açaí	3
BGCT-Cueca virada (crostoli)	2

Figura 5: Primeiras receitas com a palavra *açúcar* nos ingredientes

três diferentes partes da receita. (Por exemplo, no caso da torta de banana pacová, no BGCT, aparece como “Canela e açúcar para polvilhar”, “12 colheres (sopa) de açúcar”, e “1 xícara de açúcar”, que são, naturalmente, para serem usados em diferentes etapas da receita.)

6. Em jeito de conclusão

Neste artigo o nosso objetivo primordial foi apresentar o CulBras, um novo recurso disponível a todos os interessados em culinária, em tradução, e particularmente na culinária brasileira, no par de línguas português e inglês. Ao mesmo tempo,

também produzimos algumas pistas de como usar o corpus, e argumentamos sobre o interesse de estudar especificamente a tradução da culinária, apontando para muitos estudos já nesta área, esperando despertar o apetite de mais utilizadores para o tema.

Esperamos que num futuro próximo este corpus possa ser usado especificamente na formação de tradutores, e que possa levar à criação de novos corpora culinários, muito particularmente incluindo a linguagem culinária em Portugal.

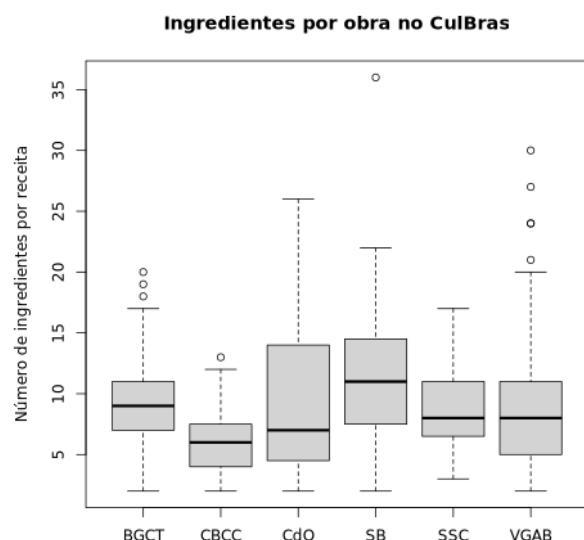


Figura 6: Número de ingredientes por receita, por obra

Agradecimentos

As três autoras agradecem-se mutuamente, porque sem a colaboração de todas o CulBras não existiria publicamente acessível no CorTrad. Em relação ao CorTrad, agradecemos a Elisa Duarte Teixeira pela sua participação ativa na criação do CpA, e a Heloisa Bacellar por ter dado autorização para o CpA fazer parte do CorTrad.

Agradecemos também sinceramente à Elisa Duarte Teixeira pela leitura cuidadosa de uma primeira versão.

A primeira autora agradece à FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pela bolsa recebida para realização do doutorado, bem como aos bolsistas de Iniciação Científica do projeto Culinária para Fins Acadêmicos (Universidade Federal do Rio Grande do Sul/UFRGS) pela revisão do alinhamento das receitas.

A segunda autora agradece à FCCN e à FCT o apoio à Linguatca ao longo dos tempos, em particular o alojamento presente dos servidores, assim como ao Research Computing Services of Sigma, Norway, pela utilização de HPC (projeto nn9186k).

Referências

Bacellar, Heloísa. 2008. *Cozinhando para amigos*. DBA. Cooking for friends (2005). Tradução de Elisa Duarte Teixeira e Alvar H. de Campos Andrade Lamparelli; revisão de Virgínia Klie

Bick, Eckhard. 2000. *The parsing system Palavras: Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus University. Tese de Doutorado

Bick, Eckhard. 2014. PALAVRAS, a constraint grammar-based parsing system for Portuguese. Em Tony Berber Sardinha & Thelma de Lurdes São Bento Ferreira (eds.), *Working with Portuguese Corpora*, 279–302

Brown, Peter., John Cocke, Stephen Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer & Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics* 16(2). 79–85. [↗](#)

Chiaro, Delia & Linda Rossato. 2015. Food and translation, translation and food. *The translator* 21(3). 237–243. [doi 10.1080/13556509.2015.1110934](#)

Evert, Stephanie & The CWB Development Team. 2022. *The IMS Open Corpus Workbench (CWB) CQP interface and query language manual*—CWB version 3.5. [↗](#)

Frankenberg-Garcia, Ana & Diana Santos. 2002. COMPARA, um corpus português-inglês na web. *Cadernos de Tradução* 9. 61–79

Gaspari, Federico. 2015. Exploring Expo Milano 2015: A cross-linguistic comparison of food-related phraseology in translation using a comparalled corpus approach. *The translator* 21(3). 327–349. [doi 10.1080/13556509.2015.1103099](#)

Ghafarian, M., Reza Kafipour & Afshin Soori. 2016. Domestication and foreignisation strategies in restaurant menu translation. *Pertanika* 24(4). 1417–1429

Grefenstette, Gregory & Pasi Tapanainen. 1994. What is a word, what is a sentence? problems of tokenization. Em *3rd International Conference on Computational Lexicography (COMPLEX)*, 79–87

Grønn, Atle & Irena Marijanovic. 2010. Russian in contrast: form, meaning and parallel corpora. *Oslo Studies in Languages* 2(1). 1–24. [doi 10.5617/osla.151](#)

Haider, Ahmad S., Omair Alzghoul & Yousef Hamdan. 2023. Creating and experimenting a new parallel corpus of English-Arabic subtitles of culinary shows: A useful guide for translating food across cultures. *World Journal of English Language* 13(8). 358–358. [↗](#)

- Isabelle, Pierre, Marc Dymetman, George Foster, Jean-Marc Jutras, Elliot Macklovitch, François Perrault, Xiaobo Ren & Michel Simard. 1993. Translation analysis and translation automation. Em *5th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI)*, 201–217. 
- Johansson, Stig. 2007. *Seeing through multilingual corpora: On the use of corpora in contrastive studies*. John Benjamins.  [10.1075/sc1.26](https://doi.org/10.1075/sc1.26)
- Koehn, Philipp. 2002. *EuroParl: a parallel corpus for evaluation of machine translation*. 
- Labov, William. 1973. *Sociolinguistic patterns*. University of Pennsylvania Press
- Mayer, Thomas & Michael Cysouw. 2014. Creating a massively parallel Bible corpus. Em *9th International Conference on Language Resources and Evaluation (LREC)*, 3158–3163. 
- Oliveira, Joacyr Tupinambás. 2019. *Um corpus de aprendizes de tradução: compilação, tratamento e aplicação ao ensino por meio de planilha eletrônica*: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. Tese de Doutorado. 
- Rayson, Paul & Roger Garside. 1998. The CLAWS web tagger. *International Computer Archive of Modern and Medieval English (ICAME) Journal* 22. 121–123. 
- Rebechi, Rozane Rodrigues. 2015. *A tradução da culinária típica brasileira para o Inglês: um estudo sob o enfoque da linguística de corpus*: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo. Tese de Doutorado. 
- Roussis, Dimitris, Vassilis Papavassiliou, Prokopis Prokopidis, Stelios Piperidis & Vassilis Katsouros. 2022. SciPar: A collection of parallel corpora from scientific abstracts. Em *13th Conference on Language Resources and Evaluation (LREC)*, 2652–2657. 
- Santos, Diana. 2002. DISPARA, a system for distributing parallel corpora on the web. Em *International Conference for Natural Language Processing in Portugal (PortAL)*, 209–218.  [10.1007/3-540-45433-0_30](https://doi.org/10.1007/3-540-45433-0_30)
- Santos, Diana. 2014. PoNTE: apontando para corpos de aprendizes de tradução avançados. *Linguamática* 6(1). 69–86. 
- Santos, Diana. 2019. PANTERA: a parallel corpus to study translation between Portuguese and Norwegian. *Bergen Language and Linguistics Studies (BeLLS)* 10(1).  [10.15845/bells.v10i1.1372](https://doi.org/10.15845/bells.v10i1.1372)
- Santos, Diana, Luís Costa & Paulo Rocha. 2003. Cooperatively evaluating Portuguese morphology. Em *Computational Processing of the Portuguese Language (PROPOR)*, 259–266. 
- Santos, Diana, Stella E. O. Tagnin & Elisa Duarte Teixeira. 2011. Colours, clothes and food in CorTrad: why corpus-based translation studies are revealing. Work presented at International Computer Archive of Modern and Medieval English (ICAME). 
- Santos, Diana, Stella E. O. Tagnin & Elisa Duarte Teixeira. 2012. CorTrad and Portuguese–English translation studies: investigating colors. *Studies in Variation, Contacts and Change in English* 12. 
- Schmid, Helmut. 1994. Probabilistic part-of-speech tagging using decision trees. Em *International Conference on New Methods in Language Processing (NeMLaP)*, 
- Scott, Mike. 2012. Wordsmith tools version 6.0. Oxford University Press
- Tagnin, Stella E. O. 2003. COMET - a multilingual corpus for teaching and translation. Em *Practical Applications in Language Corpora (PALC)*, 535–540
- Tagnin, Stella E. O., Rozane Rodrigues Rebechi & Elisa Duarte Teixeira. 2022. A fraseologia das receitas culinárias – com destaque para as brasileiras. Em Ariel Novodvorski & Cleci R. C. R. Bevilacqua (eds.), *Fraseologia: enfoques contrastivos e especializados*, 441–473. EDUFU.  [10.14393/EDUFU-978-65-5824-020-4](https://doi.org/10.14393/EDUFU-978-65-5824-020-4)
- Tagnin, Stella E. O. & Elisa Duarte Teixeira. 2004. Linguística de corpus e tradução técnica - relato da montagem de um corpus multivarietal de culinária. *TradTerm* 10. 313–358.  [10.11606/issn.2317-9511.tradterm.2004.471843](https://doi.org/10.11606/issn.2317-9511.tradterm.2004.471843)
- Tagnin, Stella E.O. 2013. *O jeito que a gente diz: expressões convencionais e idiomáticas*. Disal
- Teixeira, Elisa Duarte & Stella E. O. Tagnin. 2008. *Vocabulário para culinária. Inglês–Português*. SBS
- Tiedemann, Jörg. 2009. News from OPUS - a collection of multilingual parallel corpora with tools and interfaces. Em *Recent Advances in Natural Language Processing (RANLP)*, 237–248.  [10.1075/cilt.309.19tie](https://doi.org/10.1075/cilt.309.19tie)

- Venuti, Lawrence. 2008. *The translator's invisibility: A history of translation*. Routledge
- Viana, Vander & Stella Tagnin. 2015. *Corpora na tradução*. HUB
- Véronis, Jean (ed.). 2000. *Parallel text processing*. Kluwer Academic Publishers