

A Lingüística de *Corpus* e as súas aplicacións na Fraseoloxía – un panorama de pescudas desenvolvidas na Universidade de São Paulo¹

Corpus Linguistics and its Applications in Phraseology: An Overview of Research Developed at Universidade de Sao Paulo

Stella E. O. Tagnin

Universidade de São Paulo

<https://orcid.org/0000-0002-5517-2710>

seotagni@usp.br

Resumo: Este artigo, como xa esclarece o seu título, pretende presentar un panorama das pescudas que engloban a Lingüística de *Corpus* desenvolvidas na Universidade de São Paulo, baixo a miña orientación. Logo dunha breve introdución sobre esa metodoloxía, analizo os traballos realizados nas áreas de Ensino e Aprendizaxe, e Tradución, incluíndo aí a terminoloxía bilingüe, a tradución literaria e a tradución audiovisual. Na maioría delas o foco é a fraseoloxía, xa sexa para a confección de glosarios ou para o ensino de colocacións. Tamén se presenta o Proxecto CoMET, composto por tres *corpus*: o CorTec (*Corpus* Técnico), o CorTrad (*Corpus* de Traducións) e o CoMAprend (*Corpus* de Aprendices). O artigo conclúe con comentarios sobre outras aplicacións da Lingüística de *Corpus*.

Palabras clave: Lingüística de *Corpus*, Fraseoloxía, Terminoloxía, Tradución, Ensino e Aprendizaxe.

Abstract: This article, as the title already makes it clear, aims to present an overview of research involving Corpus Linguistics, developed at the University of São Paulo under my supervision. Following a brief introduction to this methodology, I discuss the research carried out in the fields of Translation, including bilingual terminology, literary translation and audiovisual translation, and Teaching and Learning. In most of these areas the focus is on phraseology, either for the building of glossaries or for the teaching of collocations. We also introduce the CoMET Project, composed of three corpora: CorTec (Technical Corpus), CorTrad (Corpus of Translations) and CoMAprend (Learner Corpus). The article concludes with comments on further applications of Corpus Linguistics.

Keywords: Corpus Linguistics, Phraseology, Terminology, Translation, Teaching and Learning.

¹ Data de recepción: 19.11.19. Data de aceptación: 03.03.20.
Tradución feita por Raquel Sánchez González.

1. Introducción

Nunca tiven o privilexio de ser alumna da profesora Maria Aparecida Barbosa, pero os nosos camiños cruzáronse constantemente por traballarmos en áreas afíns. Varios dos meus alumnos pasaron polas mans desa gran mestra, traendo unha rica bagaxe nas áreas de Lexicografía e de Terminoloxía. Cando chegaron onda min, iniciáronse na Lingüística de *Corpus*, área á que me dedico desde 1999, cando asistimos ao primeiro curso desa disciplina coordinado polo profesor Göran Kjellmer, da Universidade de Gotemburgo.

2. A Lingüística de *Corpus*

A Lingüística de *Corpus* (LC) é un enfoque metodolóxico que analiza grandes cantidades de textos mediante ferramentas informáticas (Mc Enery, Xiao e Tono, 2006; McEnery e Hardie, 2011) Para iso, os textos deben estar en formato electrónico; ao seu conxunto denominámoslle *corpus*. A compilación dos textos débese facer seguindo un criterio, atendendo o obxectivo da pescuda. Hai dous tipos principais de *corpus*: *corpus* monolingües e *corpus* bilingües (ou multilingües), que a súa vez poden ser comparables ou paralelos. Os bilingües comparables comprenden *corpus* con textos orixinais en ambas linguas; os bilingües paralelos compóñense por textos orixinais e as súas respectivas traducións. Metodoloxicamente, a análise comeza cunha Lista de Palabras completa do *corpus* en estudo por un *software* de análise lexical; entre as máis usadas están o WordSmith Tools (Scott, 2006) e o AntConc (Anthony, s/d). O segundo paso é elaborar unha Lista de Palabras Clave, que se obtén pola comparación do *corpus* de estudo cun *corpus* de referencia. Ese *corpus* de referencia debe axeitarse á pescuda en curso (Tagnin, 2013). Os *softwares* de análise lexical posúen ferramentas que se prestan especificamente para a identificación de colocacións, xa sexa mediante a análise de liñas de concordancia, ou pola ferramenta específica denominada «collocates». A gran maioría dos traballos desenvolvidos na USP, baixo a miña orientación, dedicáronse ao estudo da convencionalidade na lingua, abordando principalmente as colocacións, xa sexa na lingua xeral ou nas linguas de especialidade, cando constitúen termos.

3. Panorama das buscas con LC na USP

Grosso modo, xa se defenderon ao redor de 30 traballos de posgraduado², entre mestrados e doutoramentos.

Área de busca	Mestrado	Doutoramento
Tradución/Terminoloxía/Fraseoloxía	11	4
Tradución/Estudos culturais	1<	3
Tradución Literaria	-	3
Tradución Audiovisual	1	-

² A maioría dos traballos citados están dispoñibles no Banco da Datos Bibliográficos da USP – Dedalus (www.dedalus.usp.br).

Tradución/Desenvolvemento de <i>software</i>	1	-
Tradución/Interpretación	1	-
Ensino e Aprendizaxe	7	4
Total	22	14

Como se ve arriba, coa excepción do Ensino e Aprendizaxe, todas as outras áreas están, dalgún xeito, relacionadas coa tradución unha vez que son estudos bilingües e, en gran parte, tamén coa terminoloxía, con foco na fraseoloxía.

3.1. Tradución/Terminoloxía/Fraseoloxía

Ao ser impracticable, neste artigo, contemplar todos os traballos que tratan algún aspecto fraseolóxico, fíxose a selección perseguindo o obxectivo de exemplificar as áreas mencionadas arriba e salientar o impacto que cada un dos traballos pode ter. Na área de Tradución/Terminoloxía/Fraseoloxía, varios traballos estudaron os termos de diversos dominios con vistas á creación dunha obra terminográfica; por exemplo, en Medicina, termos específicos da Hipertensión Arterial (Castanho, 2003), dos Negocios (Orenha, 2004), do Ecoturismo (Martins, 2007), da Gastronomía (Teixeira, 2004; 2008; Rebechi, 2011; 2015), da Hostalaría (Navarro, 2012), do Turismo (Marqueze, 2018), entre outros. A maioría deles baseouse nunha listaxe a partir de *corpus* comparables, o que garante a autenticidade de uso dos termos recompilados. Exclúense os traballos de Rebechi (2015) e Marqueze (2018), que traballaron tanto con *corpus* comparables como con paralelos, construindo vocabularios da Gastronomía Brasileira e das Festas Folclóricas Brasileiras, respectivamente, con informacións enciclopédicas e ilustracións cando fosen pertinentes, ademais dos correspondentes nas dúas linguas, inglés e portugués.

Entre os traballos que tenden máis especificamente á fraseoloxía, destaco *A tradução de binômios nos contratos de Common Law à luz da Linguística de Corpus* (Fonseca, 2007), *Cafés do Brasil - Estudo de variantes em português e inglês na língua falada* (Ginezi, 2008), *O tradutor pode estar redondamente enganado: um estudo contrastivo de colocações adverbiais (inglês-português) sob o enfoque da Linguística de Corpus* (Moraes, 2005), *O advérbio em fraseologias das áreas da culinária e do direito contratual: um estudo baseado em corpus* (Moraes, 2011) e *A fraseologia do futebol: um estudo bilíngüe português-inglês direcionado pelo corpus* (Matuda, 2011).

3.2. Tradución/Estudos culturais

Ao analizar os verbos que acompañaban o substantivo ‘gol’ en portugués e ‘goal’ en inglés, o estudo de Matuda revelou que, mentres que en portugués só hai practicamente dous verbos, ‘*marcar um gol*’ [en galego “marcar un gol”] e ‘*fazer um gol*’ [en galego “meter un gol”], en inglés hai unha profusión que indica, inclusive, o modo no que se fai o gol (por exemplo, *slide the ball home* indica con que facilidade se marcou o punto). Tamén revelou que unha boa parte deses verbos denotan forza, velocidade e agresividade, como *thundered a ball home* (= explosión) ou *smash a ball home* (= acción violenta). Esas descubertas levaron á autora a buscar as diferenzas entre a práctica do deporte en Brasil e en Gran Bretaña, dando como resultado a súa tese *Futebóis: uma análise*

do léxico do fútbol en portugués brasileiro e inglés británico sob a perspectiva da lingüística de Corpus (Matuda, 2015).

Logo, buscando diferenzas culturais, Fonseca (2014), a partir dun corpus con relatos de mulleres brasileiras e americanas que deron a luz de xeito normal despois dunha cesárea, concluíu que estas mulleres son inducidas polos seus médicos a que opten por unha cesárea, aínda que se dean as condicións adecuadas para ter un parto natural. Outro traballo que detectou claras diferenzas culturais con base nun *corpus* de recensións de hoteis no Brasil e nos Estados Unidos foi o de Navarro Fuchs (2018), quen comprobou que a cultura brasileira era unha cultura *high context*, é dicir, privilexia un estilo de comunicación máis económico, de xeito que gran parte da mensaxe pode ser inferida do contexto; por outra banda, a cultura estadounidense enténdese como *low context*, é dicir, adopta preferentemente un xeito de comunicación máis detallado, buscando clarificar máis os elementos contextuais. Esas diferenzas débense ter en conta á hora de facer a tradución dunha páxina web dun hotel, por exemplo. Noutras palabras, unha páxina web dun hotel brasileiro que procure achegarse ao público estadounidense debe ofrecer moitos máis detalles das súas facilidades que a versión en portugués, xa que os brasileiros presentan unha menor tolerancia ao exceso de información.

Nunha perspectiva da Análise Crítica do Discurso, Broz (2019) investiga a evolución do uso de colocacións co termo *Brexit*, como *hard Brexit* e *soft Brexit*. O seu *corpus* de análise provén de xornais e tabloides da prensa británica extraídos do *corpus* NOW (News on the Web) (<https://www.english-corpora.org/now/>).

3.3. Tradución literaria

Malia aínda estar pouco difundida para a análise de obras literarias, a Lingüística de *Corpus* préstase para eses estudos, como comprobaron as teses de doutoramento de Gonçalves (2006), Aguiar (2010) e Ramos (2018). Gonçalves estudou os contos de James Joyce en *Dubliners*. Contrastando un *corpus* de referencia composto por obras de autores ingleses contemporáneos a Joyce co seu *corpus* de estudo composto polos 15 contos que integran *Dubliners*, a autora detectou peculiaridades da obra dese autor irlandés, tais como unha temática urbana en contraste coa predominancia de temas rurais nos autores ingleses. Ao analizar os verbos que acompañaban as figuras femininas nos contos, puido confirmar unha crenza enraizada na crítica literaria de que as personaxes femininas nesas obras son as máis submisas das submisas. Esa confirmación foi obxectivamente baseada no feito de que na maioría dos casos os verbos eran axentes, indicando actitudes positivas, unha iniciativa por parte das mulleres en cuestión.

Aguiar (2010) analizou as traducións ao inglés de tres novelas de Chico Buarque, cada unha feita por un tradutor diferente. A partir de *corpus* paralelos, compostos por cada novela e cadansúa tradución, o seu obxectivo era verificar como a linguaxe creativa de Chico Buarque se recupera nas traducións. A lista de palabras clave das novelas evidenciou un campo semántico —partes do corpo— do cal o propio escritor non tiña consciencia, conforme llo confesou ao doutorando. Buscando os contextos das aparicións das palabras dese campo, como *olhos*, *boca*, *língua* etc. e as súas respectivas traducións,

o autor discutiu as escollas dos tradutores. Outro traballo que analizou a tradución ao inglés dunha obra dun escritor brasileiro foi o de Silva (2015), quen investigou como os xogos de palabras na novela *O Xangô de Baker Street*, de Jô Soares, foron —ou non— recuperados na tradución. Ramos (2018) propúxose verificar se a novela *Huckleberry Finn* era, de certo, unha novela racista, segundo o divulgou a crítica literaria —ata o punto de ser prohibido nos programas escolares nos Estados Unidos—. Para rebater esa visión, a autora compilou un *corpus* con narracións de escravos co fin de verificar o uso da palabra *nigger*. Concluíu que o vocábulo era de uso común na época, sen ningunha connotación racial.

3.4. Tradución audiovisual

Unha área en crecente expansión é a da Tradución Audiovisual. Nese campo, o traballo de Nascimento (2018) destaca por centrarse en como os efectos sonoros se traducen nos subtítulos para xordos e xente con problemas de audición. A autora compilou un corpus comparable con subtítulos orixinais de películas en portugués, inglés e francés. Observou que non hai ningunha normativización na descrición deses sons, o que dificulta a comprensión do xordo/persoa con perda auditiva, xa que debe descodificar cada subtítulo individualmente. A autora propón unha *convencionalización* para cada son estudado, co propósito de facilitar e axilizar a comprensión do seu público de destino.

3.5. Tradución/Desenvolvemento de *software*

A construción dun *Vocabulario Técnico Online* para aprendices de Tradución, denominado *VoTec* foi o proxecto de Fromm (2008). Trátase dunha ferramenta que se nutre de *corpus* técnicos para a construción de entradas e dun banco de datos para o seu funcionamento. A innovación proposta supón a posibilidade de customización dos seus modos de presentación en pantalla, posibilitando dous xeitos de visualización da microestrutura: a normal, co formato dos dicionarios impresos, e a descritiva, un xeito propio dos dicionarios electrónicos, nos que a microestrutura se describe de forma xerárquica.

3.6. Ensino e aprendizaxe

Na área do Ensino, a Lingüística de *Corpus* móstrase extremadamente prometedora, tanto para detectar cuestións para seren traballadas cos alumnos, como para a confección de material didáctico. Tamén debemos salienta a variedade de áreas de ensino contempladas. Santos (2011) presentou unha proposta de ensino de colocacións adverbiais na área de negocios con base no *corpus* que compilou composto por periódicos e relatos de empresas británicas e americanas. Unha estratexia similar foi empregada por Andreetto (2013), que compilou artigos da área de Medicina e preparou aos seus alumnos para se presentaren a unha proba de dominio. Kindermann (2011) xa propuxera unha interface de Lingüística de *Corpus* coa área de Multiletramentos. A autora, que nesa época exercía de profesora nunha universidade privada, construíu un *corpus* de aprendices con redaccións dos seus alumnos que lle serviron de modelo para as súas aulas na graduación dos futuros profesores de lingua inglesa. Rosa (2014) tamén combinou a Lingüística de *Corpus* con outro enfoque, neste caso a Gramática de Construcións (Goldberg, 1995). O seu

obxectivo era describir o comportamento constitucional e lexical do verbo *get* [en galego “ter”, “obter”, “acadar”, “recibir”, etc.]. Para elaborar o seu *corpus* de estudo, extraeu 9210 enunciados do *Corpus of Contemporary English* (COCA). Os seus resultados demostraron 13 contextos constitucionais do verbo, pero salientaron a preferencia de *get* pola construción transitiva co sentido de ‘obtención’. No nivel lexical, catalogou 74 unidades fraseolóxicas nucleadas polo verbo *get*.

Murakami (2016) tamén se valeu dun *corpus* de aprendices para estudar os problemas relativos ao uso de colocacións verbais (verbo + substantivo, como *play the piano* [en galego “tocar o piano”], *hold a meeting* [en galego “celebrar unha reunión”], *solve a problem* [en galego “resolver un problema”], etc.). O *corpus* usado foi o EF-Cambridge Open Language Database (EFCAMDAT), desenvolvido pola Universidade de Cambridge, composto por redaccións de aprendices brasileiros de inglés cun nivel avanzado. Os seus resultados indicaron inadecuacións nun 25 % dos casos, a maioría causada pola interferencia da lingua materna como, por exemplo, **use corsets* no lugar de *wear corsets* [en galego “levar un xustillo”], **complete 71 years* en vez de *turn 71* [en galego “cumprir os 71”], **has 38 meters* por *is 38 meters* [en galego “ten 38 metros”], entre outros.

Oliveira, en *A Lingüística de Corpus na formación do tradutor: compilação e proposta de análise de um corpus paralelo de aprendizes de tradução*, dedicouse ao ensino da tradución, en primeiro lugar coa dirección inglés → portugués (Oliveira, 2014), e despois na dirección inversa (Oliveira, 2019). Para iso, construíu unha folla de cálculo electrónica para que as producións dos alumnos, separadas por períodos sintácticos, se puidesen proxectar en conxunto e, así, seren discutidas por todos os alumnos. Deses traballos xurdiron dous *corpus* de aprendices, un coas traducións para a lingua materna e outro coas traducións para a lingua estranxeira. Dese material, o autor extraeu os modelos para as súas aulas, é dicir, puido centrarse especificamente nos problemas fraseolóxicos dos alumnos, entre outros que xurdiron, en vez de pautarse por tópicos preestablecidos.

Unha pescuda que puido ter impacto no currículo da tradución, en especial no que respecta ao ensino de interpretación, foi a de Ginezi (2015) que construíu un *Corpus* de Ensino de Interpretación Simultánea co fin de pescudar se o ensino da interpretación consecutiva debe, de certo, ser un pre-requisito para a aprendizaxe da interpretación simultánea, como ocorre en todos os cursos. Os seus datos, analizados baixo a luz da Lingüística de *Corpus* coa axuda de entrevistas retrospectivas cos alumnos que participaron na pescuda, revelaron que as dúas modalidades esixen procesos cognitivos distintos, de xeito que poden ensinársese paralelamente. Digno de mencionar foi o feito de que as incorreccións fraseolóxicas ocorresen tanto na interpretación consecutiva como na simultánea.

Un *corpus* de lingua oral serviu como base para os traballos de Prado (2015; 2019). O seu obxectivo era a preparación de pilotos para a proba de dominio en inglés aeronáutico. Eses profesionais deben acadar o nivel 4 para teren permiso de realizar voos internacionais. O exame avalía as seguintes habilidades: Pronuncia, Estrutura, Vocabulario, Fluidez, Comprensión e Interacción. No primeiro traballo, Prado analizou, a partir dun *corpus* de diálogos auténticos entre pilotos e controladores aéreos en situacións non rutineiras, as

estruturas lexicogramaticais máis recorrentes. Os resultados demostraron que a linguaxe — estruturas e vocabulario— usada por eses profesionais non corresponde aos tópicos suxeridos pola OACI (Organización de Aviación Civil Internacional), nin ao contido programático dos materiais didácticos utilizados na época. No segundo traballo, Prado (2019) dedicouse ao estudo das habilidades de fluidez e interacción, a partir dunha visión pragmática, tendo de novo como base o *corpus* dos diálogos mencionados anteriormente. De novo, os seus descubrimentos demostraron unha descompensación entre o que e como se ensina e a realidade das interaccións entre pilotos e controladores aéreos. Nesa mesma liña, Tosqui-Lucks (2019) desenvolveu actividades didácticas a partir de *corpus* de comunicacións piloto-controlador en situacións rutineiras e non rutineiras en lingua inglesa, material que ben pode ser usado en aulas presenciais, ben a distancia en cursos para controladores aéreos. Considerando que faltas de comprensión entre eses profesionais poden, co tempo, causar incidentes ou accidentes, débese considerar a relevancia destes estudos.

4. O Proxecto CoMET

Non podería deixar de mencionar o proxecto CoMET (www.comet.fflch.usp.br), que se compón por tres *corpus* distintos: o CorTec, o CorTrad e o CoMAprend, todos eles con acceso gratuito. O CorTec é un *corpus* técnico, composto por *corpus* comparables en máis de 20 dominios, moitos dos cales conteñen os *corpus* construídos polos alumnos para as súas pescudas. O CorTrad é un *corpus* paralelo, composto por textos xornalísticos, literarios e técnico-científicos orixinais coas súas respectivas traducións. O idioma orixinal dos textos xornalísticos e técnico-científicos é o portugués con traducións ao inglés, mentres que o idioma orixinal dos literarios é o inglés con traducións ao portugués. Eses dous *corpus* permiten varios tipos de pescudas, en especial formas convencionais de expresarse, no que se inclúen os diversos tipos de fraseoloxismos. A páxina web tamén dispón de varias publicacións e instrucións de ferramentas usadas para as buscas con *corpus*.

O CoMAprend é un *corpus* de aprendices composto por redaccións de alumnos de cursos de grao en Letras nas cinco linguas ofrecidas polo Departamento de Letras Modernas: alemán, español, francés, inglés e italiano. Está aberto a contribucións de outras universidades brasileiras. Polo momento, a maior parte do material dispoñible é de aprendices da lingua inglesa, polo que a páxina web pasa actualmente por unha reforma total.

5. Consideracións finais

Como se entende nas pescudas aquí listadas, a Lingüística de *Corpus* (como metodoloxía) facilita a busca que engloba unha gran parte de datos e pode servir de apoio a varios enfoques teóricos, como a Fraseoloxía, a Terminoloxía, os Estudos da Tradución, os Estudos Culturais e, en especial, a diversas teorías do Ensino e Aprendizaxe. En todas esas áreas, o foco dos traballos recaeu sobre algún aspecto fraseolóxico, xeralmente sobre

as colocacións, xa sexan da lingua xeral, como no caso do ensino da lingua inglesa, ou na lingua de especialidade, como no caso da compilación de glosarios terminolóxicos. Os *corpus* procedentes dos traballos presentados integran o CorTec, cando son comparables, e o CorTrad, cando son paralelos, ambos incluídos no Proxecto CoMET, desenvolvido na nosa universidade.

Malia que as pescudas aquí expostas teñen esencialmente un foco lingüístico, máis especificamente lexical e fraseolóxico, a Lingüística de *Corpus*, de certo, préstase a estudos noutros eidos (Baker, 2006; Lu 2015; Stuart, Botella e Ferri, 2016) que poden ser explorados por aqueles lectores que desexen aventurarse en novos desafíos.

6. Referencias bibliográficas

- AGUIAR, S. M. (2010): *As vozes de Chico Buarque em inglês: tradução e linguística de corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- ANDREETTO, M. (2013): *A Linguística de Corpus nos cursos preparatórios para exames de proficiência*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- ANTHONY, L. (s/d): *AntConc Homepage*. Acceso o 9 de xuño de 2020, dispoñible en <http://www.laurenceanthony.net/software/antconc/>
- BAKER, P. (2006): *Using Corpora in Discourse Analysis*. Londres - Nova Iorque: Continuum.
- BROZ, V. (2019): *A corpus-based critical discourse analysis of Brexit in the English language press*. Investigación posdoutoral. San Paulo, SP: Universidade de San Paulo.
- CASTANHO, R. M. (2003): *Proposta para a Elaboração de um Glossário de Colocações na Área Médica - Subárea Hipertensão Arterial*. Disertación de Mestrado. San Paulo, SP: Universidade de San Paulo.
- FONSECA, L. C. (2007). *A tradução de binômios nos contratos de common law à luz da lingüística de corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2014): *“Eu quero cesárea!” ou “Just cut it out!”: Análise Crítica do Discurso de relatos de parto normal após cesárea de mulheres brasileiras e estadunidenses à luz da Lingüística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- FROMM, G. (2008): *VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução*. Tese de doutoramento. San Paulo: Universidade de San Paulo.
- GINEZI, L. L. (2008): *Cafês do Brasil - Estudo de variantes em português e inglês na língua falada*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2015): *Ensino de interpretação simultânea na graduação: uma análise de corpora de aprendizes*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- GOLDBERG, A. (1995): *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press.

- GONÇALVES, L. B. (2006): *Dubliners sob a lupa da lingüística de corpus: uma contribuição para a análise e a avaliação da tradução literária*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- KINDERMANN, C. A. (2011): *Nova interface pedagógica: lingüística de corpus + multiletramentos na formação do professor de língua inglesa*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- LIU, B. (2015): *Sentiment Analysis: mining sentiments, opinions, and emotions*. Cambridge: Cambridge University Press.
- MARQUEZE, G. M. (2018): *Proposta de um vocabulário bilingue de festas populares brasileiras baseada em um estudo de corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- MARTINS, J. C. (2007): *A terminologia do Ecoturismo como espelho de diferentes visões: agências de turismo, ambientalistas e governo*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- MATUDA, S. (2011): *A fraseologia do futebol: um estudo bilingue português-inglês direcionado pelo corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2015): *Futebóis: uma análise do léxico do futebol em português brasileiro e inglês britânico sob a perspectiva da lingüística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- MCERNERY, T. e HARDIE, A. (2011): *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
- MCERNERY, T.; XIAO, R. e TONO, Y. (2006): *Corpus-Based Language Studies. An advanced resource book*. Londres - Nova Iorque: Routledge.
- MORAES, H. F. (2005): *O tradutor pode estar redondamente enganado: um estudo contrastivo de colocações adverbiais (inglês-português) sob o enfoque da Lingüística de Corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2011): *O advérbio em fraseologias das áreas da culinária e do direito contratual: um estudo baseado em corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- MURAKAMI, D. S. (2016): *Colocações verbais em um corpus de aprendizes brasileiros de inglês*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- NASCIMENTO, A. K. (2018): *Convencionalidade nas legendas de efeitos sonoros na legendagem para surdos e ensurdecidos (LSE)*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- NAVARRO FUCHS, S. (2018): *Orientações culturais e suas implicações para a tradução funcionalista: um estudo na área do turismo à luz da Lingüística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- NAVARRO, S. M. (2012): *Glossário bilingue de colocações de hotelaria: um modelo à luz da Lingüística de Corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.

- OLIVEIRA, J. T. (2014): *A Linguística de Corpus na formación do tradutor: compilación e proposta de análise de un corpus paralelo de aprendizes de tradución*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2019): *Um corpus de aprendizes de tradución: compilación, tratamento e aplicación ao ensino por meio de planilha eletrônica*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- ORENHA, A. (2004): *A compilación de um glossário bilingüe de colocações, na área de negócios, baseado em corpus comparável*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- PRADO, M. C. (2015): *Levantamento dos padrões léxico-gramaticais do inglês para aviação: um estudo vetorado pela Linguística de Corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2019): *A relevância da pragmática no ensino do inglês aeronáutico: um estudo baseado em corpora*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- RAMOS, V. L. (2018): *Será Huckleberry Finn mesmo um romance racista? - Uma análise da obra, de algumas de suas traduções e do discurso racial no século XIX em narrativas sobre escravos sob a luz da Linguística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- REBECHI, R. R. (2011): *A imagem do Brasil através da sua culinária: contribuições terminológicas*. Disertación de Mestrado. San Paulo, SP: Universidade de San Paulo.
- (2015): *A tradução da culinária típica brasileira para o inglês: um estudo sob o enfoque da Linguística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- ROSA, R. G. (2014): *Fraseologia do verbo get na língua inglesa: uma abordagem da Linguística de Corpus e da Gramática das Construções*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- SANTOS, A. G. (2011): *Working closely with corpora - Proposta de ensino de colocações adverbiais em inglês para negócios, sob a luz da Linguística de Corpus*. Disertación de mestrado. San Paulo, SP: Universidade de San Paulo.
- SCOTT, M. (2006): *WordSmith Tools*. Acceso o 10 de xaneiro de 2020, dispoñible en <http://www.lexically.net/wordsmith/downloads/>
- SILVA, N. R. (2015): *A tradução de jogos de palavras no romance “O Xangô de Baker Street”: uma revisão do quadro de estratégias de Delabastita com o auxílio da Linguística de Corpus*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- STUART, K.; BOTELLA, A. e FERRI, I. (2016): *A Corpus-Driven Approach to Sentiment Analysis of Patient Narratives*. In: A. O.-H. (ed.), *CILC 2016, 1*, pp. 381-395.
- TAGNIN, S. E. (2013): *O jeito que a gente diz: combinações consagradas em inglês e português*. San Paulo: Disal.

- TEIXEIRA, E. D. (2004): *Receita qualquer um traduz - será : a culinária como área técnica de tradução*. Dissertação de mestrado. San Paulo, SP: Universidade de San Paulo.
- (2008): *A linguística de corpus a serviço do tradutor: proposta de um dicionário de Culinária voltado para a produção textual*. Tese de doutoramento. San Paulo, SP: Universidade de San Paulo.
- TOSQUI-LUCKS, P. (2019): *Elaboração de um módulo de atividades didáticas a partir de corpora de comunicações piloto-controlador em língua inglesa*. Trabalho posdoutoral. San Paulo, SP: Universidade de San Paulo.

