

A LINGUÍSTICA DE *CORPUS* E OS HETEROSSEMÂNTICOS NO PAR DE LÍNGUAS ESPANHOL/PORTUGUÊS

Carla Machado de Sá Stein- Professora efetiva de Espanhol do Instituto Federal Catarinense – Campus Camboriú – SC (IFC). Mestre em Educação (UNIVALI) e Doutoranda em Estudos da Tradução na Universidade Federal de Santa Catarina (UFSC).
E-mail: carla_machado@ifc-camboriu.edu.br

Resumo

O ensino/aprendizagem dos heterossemânticos, a Linguística de Corpus, o CREA e o *Corpus del Español*. Pretende-se comprovar: 1) Semelhanças e diferenças nas ocorrências dos heterossemânticos encontrados na Coleção *Enlaces* no CREA e no *Corpus del Español*; 2) Qual dos heterossemânticos apresentados na Coleção *Enlaces* teve maior ocorrência no *Corpus del Español* e no CREA?

Resumen

La enseñanza/aprendizaje de los heterosemánticos, la Linguística de *Corpus*, el CREA y el *Corpus del Español*. Se pretende comprobar: 1) Semejanzas y diferencias en las ocurrencias de los heterosemánticos en la Colección *Enlaces* en el CREA y en el *Corpus del Español*; 2) ¿Cuál de los heterosemánticos presentados en la Colección obtuvo mayor ocurrencia en el *Corpus del Español* y en el CREA?

Introdução

De acordo com Almeida Filho “O ensino de línguas tem uma história de muitos séculos, mas foi durante e no pós-guerra que ele se assumiu oficialmente como científico pela primeira vez” (2009, p.11). Ao longo dos tempos, da história e das interações entre falantes das duas línguas, portuguesas e espanholas sempre consideraram a compreensão e a comunicação na língua do outro como um trabalho de menor dificuldade, *facilitada* pela proximidade dos dois idiomas (grifo nosso). Professores de Espanhol como Língua Estrangeira (doravante E/LE) sabem da facilidade com que os alunos costumam ler, entender e comunicar-se na língua meta, porém ao mesmo tempo se adverte a produções de erros aparentemente simples, que são cometidos sistematicamente nessas comunicações.

A presente pesquisa resulta da vivência da prática docente desta pesquisadora, e tem como finalidade relatar essa experiência, ainda em andamento, enfatizando a metodologia utilizada e os ajustes feitos até o momento. Como professora de espanhol, há quatorze anos, observa-se o comportamento de alguns alunos ao fazerem uso de algumas ferramentas de tradução como: dicionários impressos/*online* ou *softwares* de tradução. Quando se trata da tradução dos heterossemânticos¹ nota-se certa influência na compreensão e na tradução de textos em espanhol para o português brasileiro e constata-se o porquê de alguns alunos optarem pelo estudo do espanhol ao invés do inglês: é que devido à proximidade lexical da Língua Espanhola com a Língua Portuguesa (doravante LP) a aprendizagem torna-se mais *fácil*, porém os causadores de equívocos no âmbito lexical, seriam os heterossemânticos (grifo nosso).

Bechara e Moure (1998) dizem que línguas muito próximas como a Espanhola e a Portuguesa apresentam um número elevado de cognatos. Vejamos alguns exemplos: *amar*

(E) /amar (P), *consideración* (E) /consideração (P), *engañar* (E) /enganar (P); “*aceitar*, do Espanhol, traduz-se por untar e lubrificar em Português, e não por aceitar. Para o nosso aceitar, o correspondente é *aceptar*. *Oso* traduzimos por urso. O termo espanhol para osso é *hueso*” (Bechara e Moure, 1998, p.11). Os autores afirmam que é necessário ter atenção nas semelhanças entre as palavras de línguas distintas, pois nem sempre isso representa *facilidade* no entendimento. “Pode, ao contrário, constituir um problema. Isso porque, ao lado dos cognatos, encontramos também, comparando o léxico de línguas diferentes, os falsos cognatos” (Bechara e Moure, 1998, p.10, grifo nosso).

A respeito da *facilidade* que brasileiros aprendizes do E/LE dizem possuir, Almeida Filho expõe que “Línguas muito próximas levam o aprendiz a viver numa zona de facilidade enganosa proporcionada pelas percepções dos aprendizes” (1995, p.19 *apud* Silva, 2002, grifo nosso).

O tema da *suposta facilidade* na aprendizagem de uma Língua Estrangeira (doravante LE) também é abordado por Vandresen (1988). O autor afirma que a *facilidade* ou dificuldade em aprender um segundo idioma está relacionada à proporção de divergência existente entre suas estruturas, dessa maneira o espanhol torna-se *fácil* para aprendizes brasileiros, porém mais difícil para os nativos da língua inglesa, alemã ou chinesa (Vandresen, 1988, p.77, grifo nosso). Durão (2004, p. XVIII) assegura que não é o mesmo ensinar espanhol para falantes de uma língua tão próxima como é a portuguesa que fazê-lo a outros falantes de línguas mais distantes como o alemão, o árabe e o japonês. A autora expõe que os falantes de línguas próximas cometem muitos erros, talvez não menos importante que os de línguas distantes, mas esses erros não recaem com tanta frequência em estruturas decisivas para a comunicação e, apesar de sua abundância, a dificulta em menor escala. Durão (2004) diz que a proximidade ou a semelhança entre dois idiomas *facilita* apenas quando do início da aprendizagem (Durão, 2004, p.38, grifo nosso).

Segundo Ulsh (*in* Krebs, 2007), devido à ‘semelhança’ entre a Língua Espanhola e a LP nos primeiros dias de aula, o aluno brasileiro que opta pelo E/LE apresenta certa fluência na língua em estudo. Ulsh diz que tal situação proporciona a ilusão de que falar Espanhol é fácil e o aprendiz se considera falante de Espanhol, porém essa falsa facilidade faz com que haja a prática do “portunhol” (mistura do Português e do Espanhol), ou seja, ocorre uma interferência da Língua Materna (doravante LM) na Língua-alvo. O “portunhol” geralmente ocorre quando o aprendiz procura primeiro adquirir a língua por imersão, tentando se comunicar em espanhol com as pessoas que estão a sua volta, colegas de estudo ou de trabalho, etc.

De acordo com Bechara e Moure: “Mesclar as duas línguas é prática corrente”. O conhecido “portunhol, constitui frequentemente um ‘recurso salvador’ do aprendiz de uma e de outra língua” (1998, p.9, 10). O “portunhol quebra o galho”, porém provoca uma

paralisação na produção do aluno, seu uso pode ter êxito apenas em uma necessidade imediata de comunicação. Em outras circunstâncias, onde é exigido certo desempenho do falante, não apenas é insuficiente, como acaba se transformando em um ‘inimigo’. E uma maneira que se propõe trabalhar esse “portunhol” seria o uso do *corpus*, o que de acordo com a autora Gerber (2007), em se tratando de língua, é preciso ter em mente que um *corpus* é uma amostra de uma população. “Isso não impede a eficiência de trabalhar com *corpus* linguístico, pois já se sabe o quão revelador ele pode ser, mas requer que se levem em conta as orientações da metodologia que norteia pesquisas dessa natureza” (2007, p.179, 180).

Para se trabalhar o *corpus* faz-se necessário recorrer à informática. Sabe-se que o desenvolvimento recente da tecnologia informática e da Linguística de *Corpus* (doravante LC) propicia avanços expressivos como inovação metodológica e possibilidade de consulta a uma grande quantidade de informação que antes era inacessível. Nas últimas décadas, graças ao surgimento de programas de computador capazes de analisar grandes massas de dados, tem sido viável e cada vez mais frequente o uso de *corpora* na análise linguística. Pensando nisso, esta pesquisa teve início em 2010 com vista a fomentar a pesquisa acadêmica no campo do ensino/aprendizagem do E/LE, inserida na linha de pesquisa "Lexicografia, tradução e ensino de línguas", a qual se propõe a:

contemplar as várias facetas do fenômeno tradutório relacionadas a aspectos como: a lexicografia, do ponto de vista do ensino de línguas; a produção de dicionários monolíngues, bilíngues, ‘bilingualised’; a contribuição dos estudos de corpora para a produção de dicionários alternativos e para a descrição de traduções; a prática de tradução; o estudo dos processos tradutórios (cognitivos e textuais) em sua relação com a aprendizagem de línguas estrangeiras; os recursos tecnológicos e a tradução (PGET, UFSC, 2013).

Justifica-se a escolha deste estudo o fato de esta pesquisadora ser professora de espanhol há quatorze anos (escola pública/particular) e ter percebido a grande dificuldade que os alunos apresentam na aprendizagem/tradução dos heterossemânticos do espanhol para o português/brasileiro. Pretende-se refletir sobre a tradução como ferramenta de ensino de línguas, assim como sobre o emprego de novas tecnologias utilizadas na sala de aula e por fim, apontar estratégias didáticas possíveis através do uso dos novos recursos de ensino/aprendizagem da LE. Essa pesquisa mostra também algumas possibilidades de cooperação entre a Tradução e a Linguística de Corpus.

A fim de dar continuidade à 2ª etapa deste trabalho será feito um questionário (3ª etapa), o qual será aplicado no 2º semestre/2013 nos alunos do 1º, 2º e 3º ano do Ensino Médio (doravante EM) do Curso Técnico em Hospedagem do Instituto Federal Catarinense (doravante IFC) – Campus Camboriú – SC (esta etapa será publicada em outro artigo, após a defesa desta Tese). Baseado em bibliografias que tratam do assunto, confeccionou-se esse

instrumento com a finalidade de verificar se os alunos supracitados fazem uso de material de apoio nas aulas de espanhol, quais seriam esses instrumentos e se os mesmos são adequados às necessidades desses estudantes. Esse público-alvo foi escolhido por ser esta pesquisadora professora deles há um ano e meio e continuará sendo por mais dois anos para o 1º ano e por mais um ano para o 2º ano. O 3º ano encerrou as aulas de espanhol no 1º semestre/2013, porém participará da 3ª etapa desta pesquisa (questionário).

Após a compilação dos dados do questionário será possível saber se os alunos têm o hábito de usar tradutores *online* gratuitos e/ou dicionários monolíngue/bilíngue impressos ou se simplesmente traduzem do espanhol para o português brasileiro da maneira que ‘acham’ ser a correta, praticando assim o ‘portunhol’.

Durão (1998a *apud* Andrade, 2010), quando da análise da interlíngua de estudantes universitários brasileiros de espanhol, adotou como situação inicial a produção escrita e demarcou os erros léxicos em três categorias: “a primeira centra-se no *uso incorreto de falsos amigos*; a segunda trata especificamente dos casos de transferência linguística; e a terceira analisa o que a autora denomina como casos de fronteira” (1998a *apud* Andrade, 2010, p.30, 31, grifo nosso). Ao fazer a apuração dos dados, Durão concluiu que 80% dos erros léxicos eram provenientes do uso incorreto dos Falsos Amigos² (doravante FA) e da transferência linguística.

Tendo em vista a alta porcentagem dos erros léxicos provenientes do uso incorreto dos FA, pretende-se constatar os FA (1ª etapa) que aparecem na Coleção *Enlaces español para jóvenes brasileños*, depois averiguar a ocorrência deles no *Corpus del Español* e no *Corpus de Referencia del Español Actual* (doravante CREA - 2ª etapa).

Como dito anteriormente, com o objetivo de dar prosseguimento a esta tese de doutoramento, pretendemos publicar após a defesa, as demais etapas que são as seguintes: A 3ª etapa será o questionário; na 4ª etapa, os 36 alunos do 2º ano do EM - Curso Técnico em Hospedagem do IFC – Campus Camboriú – SC, irão selecionar duas ocorrências do heterossemântico que obteve o maior número de ocorrência no CREA e no *Corpus del Español*, totalizando 72 ocorrências. Dando prosseguimento, as 72 ocorrências serão traduzidas no *Google Translate* (5ª etapa). Em seguida os alunos irão retraduzir/ajustar as traduções do *Google* fazendo uso do Dicionário Escolar WMF bilíngue impresso (espanhol/português brasileiro) totalizando 144 traduções (6ª etapa).

Por perceber o destaque que a Língua Espanhola vem ganhando no Brasil e no mundo e por saber que a Internet, os *softwares* e as ferramentas de tradução exercem um grande fascínio sobre os alunos, professores de LE e profissionais da tradução, resolveu-se, então, unir o ensino da Língua Espanhola com enfoque no ensino/aprendizagem dos heterossemânticos ao uso de algumas ferramentas como: o CREA, o *Corpus del Español* (1ª e 2ª etapas; o *Google Translate* e o Dicionário WMF- 5ª e 6ª etapas).

1) A importância da Língua Espanhola

Segundo o Anuário do Instituto Cervantes (doravante CVC) hoje já falam espanhol mais de 495 milhões de pessoas, “inclusive os falantes de domínio nativo, os de domínio limitado e os falantes de espanhol como língua estrangeira. A estimativa do número de falantes para o ano de 2030 é de 535 milhões de falantes e para o 2050, de 550” (Instituto Cervantes - *El español: una lengua viva* – Informe 2012, p.7, tradução nossa).

A Lei de Diretrizes e Bases da Educação (doravante LDB, nº 9.394/96- artigo 36, inciso 3º) diz que: “será incluída *uma língua estrangeira* moderna, como disciplina *obrigatória*, escolhida pela comunidade escolar, e *uma segunda língua*, em caráter *optativo*, dentro das prioridades da instituição”. Tendo em vista a não existência da obrigatoriedade da Língua Inglesa, o espanhol ganhou um lugar de destaque e passou a ser imprescindível (grifo nosso). Ainda segundo essa lei, o ensino da Língua Espanhola é de oferta obrigatória para o centro escolar e de matrícula facultativa para o aluno. Além disso, é facultativa a inclusão da Língua Espanhola nos currículos dos quatro últimos anos do Ensino Primário.

Após a lei que obriga a oferta da Língua Espanhola nas escolas públicas de EM, estatísticas apontam que 2,4 milhões de alunos estudam espanhol no Brasil. Embora não haja dados exatos, há uma estimativa de que no Ensino Fundamental (doravante EF) a quantia de alunos pode ser ainda maior (CVC, anuário 2009). No que se refere às línguas mais faladas, a diretora do CVC (*europapress*, 2009; anuário, 2000), Carmen Caffarel, diz que, segundo revelou a revista *Ethnology* (nov/2009), o espanhol se converteu na segunda LM mais falada do mundo e é a terceira na Internet depois do inglês e do chinês.

O Brasil é o único país de fala portuguesa na América Latina, está rodeado de nações hispanofalantes, possui características econômicas, sociais e culturais semelhantes a seus vizinhos. São quase 15 mil quilômetros de fronteira dividida com nove países. Nessa faixa se localizam 588 municípios:

Do exposto se deriva que no Brasil o espanhol é, por um lado, uma língua de contato e por outro, uma língua estrangeira. O contato a que nos referimos está definido pelas características geográficas, sociais, culturais e econômicas. Nas zonas de fronteira “seca”, onde não há nenhuma barreira geográfica natural, o espanhol é uma segunda língua, como por exemplo na fronteira Brasil-Paraguai, onde as cidades se correspondem a ambos os lados da fronteira (Santanna do Livramento e Rivera) (Junior, 2010, p.1-2, tradução nossa³).

A obrigatoriedade da oferta da Língua Espanhola nas escolas públicas e privadas de EM passou a valer a partir do dia 5 de agosto de 2005 com a Lei 11.161 (conhecida como a “*Ley del Español*”) e o prazo de implantação do espanhol nas escolas públicas dado pelo governo foi de cinco anos, ou seja, o processo de implementação foi concluído em 2010. A Rede Pública de Ensino deverá ofertar a Língua Espanhola em horário regular através de centros de ensino de LE. Por sua parte, a rede particular poderá adaptar diferentes

estratégias que incluam desde classes em horário normal até cursos externos em centros de estudos de línguas modernas.

No que diz respeito ao processo de avaliação e seleção do material didático da referida língua, deu-se em outubro do mesmo ano, ficando a cargo da Secretaria de Educação Básica (doravante SEB/MEC) e do Fundo nacional de Desenvolvimento da Educação (doravante FNDE) a distribuição dos livros aos professores de espanhol do EM da Rede Pública.

Assim sendo, em 2006 as escolas que já ofereciam a Língua Espanhola receberam: uma gramática, um dicionário bilíngue espanhol/português – português/espanhol, assim como um dicionário monolíngue da Língua Espanhola. E ainda em 2006 foram publicadas as Orientações Curriculares para o EM, nas quais no vol. 1 encontra-se um item sobre os “Conhecimentos de línguas estrangeiras”, assim como um item específico para o espanhol.

Em 2008 o edital do Programa Nacional do Livro Didático (doravante PNLD, 2011) informava que o Ministério da Educação e Cultura (doravante MEC) iria adquirir livros didáticos de inglês e espanhol, desde o 6º até o 9º ano, indicando assim, a pré-inscrição das coleções de 12 de janeiro a 27 de março de 2009.

Em 2010 esta pesquisadora lecionou na Escola Municipal Prefeito Francisco Victor Alves na cidade de Itapema/SC e em 2011 no IFC Campus Avançado de Ibirama/SC e ambos não possuíam livro didático de LE para o EM. Espera-se que essa distribuição tenha sido ampliada até 2013 através do Programa Nacional do Livro Didático para o Ensino Médio (doravante PNLEM, 2012). Esta pesquisadora atualmente leciona no IFC Campus Camboriú/SC (desde 2012) e constatou que nesse campus houve a entrega dos livros didáticos de inglês e espanhol no ano de 2011.

1.1) Conceito e a classificação dos heterossemânticos, falsos cognatos ou FA na visão de alguns estudiosos da área do ensino/aprendizagem do E/LE

Através das leituras sobre os heterossemânticos, falsos cognatos ou FA, percebe-se que não há uma conceituação teórica dos termos que sejam livres de contradições. Geralmente o que se observa é que as expressões falsos cognatos e FA são consideradas sinônimas, por esse motivo são utilizadas para designarem um mesmo fenômeno linguístico (Sabino, 2006, p.251).

De acordo com Bechara e Moure (*apud* Andrade Neta, 2010) os falsos cognatos são “palavras que têm formas semelhantes ou idênticas, mas divergem quanto ao sentido” (1998, p.10).

Segundo Vita (*in* KREBS, 2007) a palavra heterossemântico começou a ser empregada no Brasil depois da obra “*Gramática para uso dos Brasileiros*” (Antenor Nascentes, 1934). A autora Krebs (2007) diz que em 1945, após a obrigatoriedade do

ensino da Língua Espanhola no Brasil (1942), surge o primeiro manual para o ensino do espanhol, o *Manual de Español*, de Idel Becker, que veio consagrar a interpretação antes feita por Antenor Nascentes (1934/1939). Andrade (2010) expõe que “é possível afirmar que a tradição contrastivista que tem como foco o binômio português-espanhol se instalou no Brasil na primeira metade do século XX, com Antenor Nascentes e Idel Becker” (2010, p.40).

Durão (2009, p.32) afirma que os heterossemânticos (“Falsos amigos” = FA, termo utilizado pela autora) determinam as unidades léxicas de duas línguas que são equivalentes ou muito semelhantes no seu formato, mas que divergem nos seus respectivos conceitos e/ou no uso que os falantes nativos fazem delas.

Lado (*apud* Silva, 2002) explica que os falsos cognatos são palavras semelhantes na forma com diferentes significados. A visão de Lado sobre os heterossemânticos é que eles:

[...] formam um grupo especial e estão situados num nível bem elevado na escala de dificuldade dessa classificação, rotulada de difícil. Tais palavras são "armadilhas infalíveis", pois a *semelhança* delas, em forma, com palavras da língua materna aumenta sua frequência no uso normal da língua pelos estudantes (1972, p.117-118, grifo nosso).

Bechara e Moure (*apud* Andrade Neta, 2010) classificam os heterossemânticos (FA de acordo com os autores) em três grupos:

- 1) Formas semelhantes com significados totalmente distintos- alguns vocábulos apresentam formas semelhantes ou idênticas nas duas línguas, porém, diferem totalmente quanto ao significado empregado nas duas línguas. Não oferecem nenhum significado comum nas duas línguas confrontadas. Por exemplo, *polvo* (espanhol) / pó, poeira (português); *pulpo* (espanhol) / polvo (português). Pode-se observar que a palavra polvo existe nas duas línguas, apresentam formas idênticas, porém, não compartilham o mesmo significado.
- 2) Formas semelhantes com um ou mais de um significado semelhante e outro ou vários distintos- os vocábulos que compõem este grupo podem compartilhar um ou vários significados, porém apresentam outro ou outros distintos que não se correspondem nas línguas confrontadas. Por exemplo, o vocábulo acordar em ambas as línguas apresenta equivalência como “estar de acordo”. Porém, vão apresentar significados muito diferentes em ambas as línguas. Em espanhol pode significar também “decidir, lembrar-se”. Em português pode significar também “despertar”. Nesses casos não há equivalência de sentido.
- 3) Formas semelhantes com significados distintos no uso atual- há vocábulos que têm uma origem comum e que chegam a compartilhar o mesmo significado em fases anteriores de ambas as línguas. Por exemplo, a palavra latir oriunda do latim que significa “pulsar, latir” (o coração) e também “ladrar” (o cachorro). Atualmente, o português mantém somente o sentido de “ladrar” enquanto que o espanhol somente mantém o significado de “bater” (Bechara e Moure *apud* Andrade Neta, 2010, tradução nossa).

1.2) A Linguística de *Corpus*

Segundo Berber Sardinha é unânime entre pesquisadores da tradução e linguistas de *corpus* o posicionamento de que “tanto os estudos tradutológicos como área acadêmica de

pesquisa, quanto à prática tradutória, têm muito a ganhar com um contato maior com a Linguística de *Corpus*” (2003, p.15).

A expressão “Linguística de *Corpus*” é nova e é cada vez maior o número de linguistas que procuram explorar a LC no estudo da linguagem. Leech (1992, *apud* Gerber, 2007, p.65) sustenta que a primeira vez que essa expressão apareceu foi em 1984 como título de um livro de Aarts e Meijs. A utilização de *corpus* sempre foi um meio empregado em pesquisas linguísticas. Gerber afirma que a LC é “uma eficiente ferramenta para estudar a língua, posto que dirige o manuseio de milhões de dados linguísticos” (2007, p.65). Murakawa (2006) cita o emprego de “*corpora* em dicionários elaborados durante os séculos XVIII e XIX, como é o caso do *Vocabulário Portuguez e Latino*, elaborado pelo Padre Rafael Bluteau e publicado entre 1712-1728, embora tenha sido concebido e realizado ainda no século XVII” (2006 *apud* Aluísio; Almeida, p.156).

De acordo com a autora Zapparoli (2009) a LC é uma ciência empírica que faz uso do computador para armazenar e analisar os fatos da língua em uso. É voltada para as investigações com base em *corporas* eletrônicos, no princípio era voltada para a área da lexicografia. Ultimamente vem crescendo o interesse: na construção de dicionários e glossários de língua e de especialidade; na área de ensino/aprendizagem de LE; nos estudos ligados à tradução; na crítica literária e nos exames dos padrões da linguagem.

O especialista em LC, Berber Sardinha, afirma que essa área vem se desenvolvendo desde a década de oitenta (séc. XX) na Europa e mais tarde em outros locais do mundo, no Brasil, a pesquisa em *corpus* ocorre em centros mais voltados para a linguística computacional, a lexicologia e o ensino de línguas. Berber Sardinha diz que a sua aplicação se dá tanto “na área da Lexicografia, quanto nos estudos sistemáticos do uso da língua, em trabalhos de Tradução, Linguística Aplicada e em Processamento de Língua Natural (doravante PLN)” (*apud* Tagnin, 2008, p.18). O autor cita Maria Tereza Biderman e Zaporolli como pioneiras no processamento da linguagem por computador no Brasil nos anos de 1960 e 1970 (*apud* Tagnin, 2008, p.34). Podemos citar o Programa de Estudos Pós-Graduados em Linguística Aplicada a Estudos da Linguagem (LAEL), da PUC/SP, por possuir um centro pioneiro em estudos da linguagem. Entre esses estudos, há também as pesquisas em *corpora* do projeto DIRECT conduzido pelo autor supracitado (1999a, 1999b, 2004).

Na área de LE temos em destaque o projeto COMET da USP liderado por Tagnin (2001), elaborado por conter material de ensino para as línguas: espanhola, italiana, inglesa, alemã e francesa, assim como traduções técnicas nas áreas da informática, direito e odontologia. Outro projeto promissor é o CORDIAL do Núcleo de Estudos da Tradução (NET) da Faculdade de Letras da UFMG, o qual está sendo desenvolvido por Alves,

Magalhães e Pagano (2001). O objetivo é o estudo de características discursivas e cognitivas por meio de uma abordagem interdisciplinar que congrega subsídios dos estudos de *corpora*, dos estudos da tradução, dos estudos da cognição, da análise do discurso e dos estudos culturais.

A LC é uma área de pesquisa que vem se consolidando nas principais universidades do Brasil. A autora Tagnin afirma que a LC “vem ganhando espaço nas universidades brasileiras, quer como disciplina em programas de pós-graduação, quer como suporte a uma variada gama de pesquisas” (2004 *apud* Gerber, 2007 p.160). Em 1998 a LC foi introduzida na Universidade de São Paulo, o prof. Göran Kjellmer (Universidade de Gotemburgo) foi o primeiro a ministrar aula da disciplina Introdução à Linguística de *Corpus* em um curso de pós-graduação (Gerber, 2007, p.166).

A LC é um tipo de metodologia de pesquisa que se destaca pelo formado trio Linguística Computacional, pela de *Corpus* e pela Estatística. Essas ferramentas se completam propiciando uma metodologia científica, em que os resultados podem ser vistos e comprovados. Com esse tipo de pesquisa pode-se demarcar “os padrões de *ocorrência* de um vocábulo (substantivo, adjetivo, artigo, preposição, pronome, etc.), de formas e tempos verbais, de determinados significados assumidos por uma expressão” (Gerber, 2007, p.46, 47 – grifo nosso).

A autora Gerber (2007, p.32) afirma que em 1960 textos de Quirk *Towards a description of English usage* (Para uma descrição do uso do inglês) foi o pontapé inicial para o seu projeto *Survey of English Usage*, o qual iniciou em 1961 abordando trabalho com *corpus*. A partir dos anos 80, com o desenvolvimento dos computadores e das ferramentas informáticas, é que foi possível selecionar, analisar, organizar e calcular dados, de modo que a possibilidade de erro na verificação dos fatos diminuiu. Por essa razão pode-se dizer que a LC e o desenvolvimento dos computadores estão unidos, uma vez que um grande número de dados precisa ser analisado de maneira breve, correta e eficiente.

Segundo Berber Sardinha, a LC é uma área que se destina à produção e análise de *corpora*, que são “conjuntos de textos e transcrições de fala armazenadas em arquivos de computador” (2009, p.7). Por esse motivo pode-se afirmar que a LC mantém uma relação direta com a tecnologia, uma vez que esta permite o armazenamento de *corpora* e sua exploração. Como lembra Berber Sardinha (2004 *apud* Gerber, 2007, p.40): “está em curso uma verdadeira revolução no pensamento linguístico, com implicações, tais como *o que é língua*, como ela é organizada, *como deve ser estudada*, *como deve ser ensinada*” (grifo nosso).

Berber Sardinha define a LC como uma área da Linguística que se incumbe da coleta e da análise de *corpora*. O autor afirma que a LC tem como função analisar “conjuntos de dados linguísticos textuais que foram coletados, criteriosamente, com o

propósito de servirem para a pesquisa de uma língua ou variedade linguística”. A LC “dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador” (2000a *apud* Alonso, 2006, p.22). Segundo Berber Sardinha, a LC “trata do uso de *corpora* computadorizados (coletâneas de textos, escritos ou de transcrições de fala, mantidas em arquivo de computador)” (2004 *apud* Alonso 2006, p.16). Uma das principais características da LC é o uso da informática para a distribuição e análise do material com finalidade de pesquisa.

De acordo com Gerber (2007, p. 78) a Linguística Computacional teve origem junto com o computador, durante a Segunda Guerra Mundial, quando os militares tiveram a necessidade de traduzir informações em russo, nascendo assim os primeiros programas de tradução automática. Com a evolução da informática, podemos armazenar e analisar um *corpus* de mais de 100 milhões de *ocorrências* (2007, p.79 – grifo nosso).

1.2.1) O que é um *corpus*

Na última década, pesquisas baseadas em *corpus* têm tido um amplo desenvolvimento no contexto brasileiro. Sua relevância e pertinência podem ser notadas na Linguística, na Linguística Aplicada e na Linguística Computacional.

Segundo Sánchez “um *corpus* é um conjunto de dados linguísticos – pertencentes ao uso oral ou escrito da língua ou a ambos – sistematizados conforme determinados critérios, suficientemente extensos em amplitude e profundidade”. (1995:8-9, traduzido por Alonso, 2006, p.23). Ou seja, tendo como objetivo o estudo linguístico um *corpus* deve compor-se por dados autênticos e legíveis por computador. Pode-se dizer então que uma pesquisa com base em *corpus* desfruta da tecnologia e da matemática para pesquisar aquilo que é incontestável no uso da língua, a partir de uma massa linguística (Gerber, 2007, p. 74). Assim sendo, conclui-se que muitos livros, ou uma grande quantidade de revistas, até mesmo textos impressos, não podem ser considerados *corpus* pela LC, pois os dados linguísticos não estão num formato que possam ser processados por um computador.

A autora Santos (2008) expõe que podemos identificar quatro tipos de usos de *corpus*:

1. Em primeiro lugar, usa-se um corpo para ter uma ideia do problema/conhecer, dando origem às metáforas do “corpo como consultor”, “corpo como familiarizador”, “corpo como treinador”, ou “corpo como mar de língua”.
2. Em segundo lugar, usa-se um corpo para medir um dado fenômeno.
3. Em terceiro e mais comumente, para avaliar algo: uma hipótese, um sistema, um método, uma teoria.
4. Finalmente, o uso talvez mais frequente é para criar outras coisas, e entre estas destaco: a) dicionários ou outras estruturas de conhecimento, como terminologias, almanaques e ontologias, b) *materiais de teste de ensino de línguas*, c) sistemas de resposta automática a

perguntas (RAP), d) sistemas de ensino, e) jogos, f) sistemas de detecção de plágio, de correio não endereçado (*spam*), ou outros (*apud* Tangin, 2008, p.51, grifo nosso).

Para resolver problemas do estudo de uma língua, pode-se fazer uma análise baseada em *corpus*, “podendo ser aplicada a vários campos da linguística aplicada, dentre os quais a *tradução*, a elaboração de dicionários e gramáticas, o *ensino de idiomas* etc” (Berber Sardinha, 2004 *apud* Alonso, 2006, p.23, grifo nosso).

De acordo com Almeida (*apud* Tagnin, 2008, p.79,80) para se chegar às fontes confiáveis de onde proverá o *corpus* é necessário realizar pesquisas, fazendo uso de motores de busca como o *Google*, por exemplo; para conhecer quais são as instituições e/ou entidades públicas, privadas ou mistas que têm tradição (grifo nosso).

Concluindo esta subseção pode-se dizer que a utilização de *corpus* faz com que haja mais praticidade no estudo da língua e obtêm-se informações efetivas sobre a língua falada e escrita (Gerber, 2007, p.61).

2) Metodologia

A metodologia utilizada neste artigo foi definida com base na leitura do material bibliográfico para a interpretação e elaboração teórica da pesquisa. Deu-se ênfase aos estudos de *corpora*, a partir dos quais sentenças em formato eletrônico são automaticamente analisadas através de programas de computador que geram dados quantitativos que são de primordial importância para o pesquisador.

Tendo conhecimento do surgimento das novas tecnologias que se incorporaram ao nosso cotidiano, à vida familiar, à escola, ao trabalho e a outros ambientes sociais, decidiu-se realizar um estudo sobre o ensino/aprendizagem dos heterossemânticos fazendo uso do CREA e do *Corpus del Español*.

De acordo com Litwin (1997):

Não se trata simplesmente da criação de tecnologia para a educação, da recepção crítica ou da incorporação das informações dos meios na escola. Trata-se de entender que se criaram novas formas de comunicação, novos estilos de trabalho, novas maneiras de ter acesso e de produzir conhecimento. Compreendê-los em toda sua dimensão nos permitirá criar boas práticas de ensino para a escola de hoje (1997, p. 131 *apud* Mendes, 2007, p.12).

O uso das ferramentas eletrônicas tem proporcionado um novo universo de pesquisas. Partindo do princípio que as ferramentas de auxílio ao ensino/aprendizagem dos heterossemânticos podem vir a colaborar com as tarefas dos tradutores, alunos e professores de LE, nos textos que contêm os heterossemânticos (do espanhol para o português brasileiro), é que este artigo, conforme especificado na introdução, tem o propósito de identificar e coletar os heterossemânticos na Coleção Enlaces - *Español para jóvenes brasileños* (1ª etapa). Em seguida verificou-se suas ocorrências no CREA e no

Corpus del Español (2ª etapa), sendo que as duas primeiras etapas no 1º semestre/2013. Os dados para essas etapas foram coletados a fim de constatar se a seleção dos FA apresentados na Coleção representam os heterossemânticos usados nas interações reais da língua.

A 1ª e 2ª etapas se guiaram pelas perguntas norteadoras da pesquisa: 1) Quais as semelhanças/diferenças de ocorrências dos heterossemânticos encontrados na Coleção Enlaces no CREA e no *Corpus del Español*? 2) Qual dos heterossemânticos apresentados na Coleção Enlaces teve maior ocorrência no *Corpus del Español* e no CREA?

Pretende-se dar continuidade a este estudo (3ª, 4ª, 5ª e 6ª etapas) após a qualificação do Doutorado, quando outro artigo será lançado. Na 3ª etapa o público-alvo desta pesquisa, ou seja, os 112 alunos de espanhol do EM do Curso Técnico em Hospedagem do IFC-Campus Camboriú/SC, responderá um questionário de 31 perguntas.

Como 4ª etapa os 36 alunos do 2º ano do EM irão selecionar 72 ocorrências encontradas do heterossemântico que obtiver o maior número de ocorrências no CREA e no *Corpus del Español*, sendo duas ocorrências por aluno, totalizando 72 ocorrências. Propõe-se analisar as 72 ocorrências em diferentes registros: língua falada, ficção, imprensa e registro acadêmico. Em seguida, essas ocorrências serão traduzidas fazendo uso do *Google Translate* (5ª etapa).

Na 6ª etapa as frases ora traduzidas no *Google Translate*⁴ serão retraduzidas/ajustadas fazendo uso do dicionário WMF, totalizando 144 traduções. Para o fim a que se destina esta pesquisa optou-se pelo *Google*, o qual é público e notório, muito utilizado principalmente por alunos e demais estudiosos de LE.

2.1) A Coleção *Enlaces español para jóvenes brasileños*

A coleção está organizada em três volumes, sendo que cada um apresenta oito unidades. O primeiro possui uma unidade zero cuja finalidade é apresentar o espanhol no mundo, os países que têm o espanhol como língua oficial, em seguida tem-se o alfabeto. As unidades estão organizadas em seções: *Competencias y habilidades*, *Funciones comunicativas*, *Contenidos lingüísticos e Género discursivo*. Ao final de cada unidade há uma autoavaliação: *Así me veo*. No término de cada volume há exercícios complementares: *Un poco más de todo*, *Vestibular* e *Te digo y me dices*. No terceiro volume esta última seção é substituída pelo Simulado para o Exame Nacional do Ensino Médio (doravante ENEM).

Em todos os volumes pode-se encontrar o *Glosario* e a *Tabla de verbos*. Cada unidade está composta didaticamente da seguinte maneira: Página de entrada, *Hablemos de...*, *¡Y no sólo esto!*, *Manos a la obra*, *En otras palabras*, *Como te decía...*, *Más cosas*, *Así me veo*.

2.2) O *Corpus del Español* e o CREA

No site do *Corpus del Español* (<http://www.corpusdelespanol.org/>), *corpus* público e gratuito, pode-se fazer buscas entre mais de 100.000.000 palavras procedentes de mais de 20.000 textos do espanhol dos séculos XIII ao XX de uma maneira rápida e simples. (Del 6)

O *Corpus* também permite fazer buscas por frequência de uso e comparar a frequência de uso de palavras, frases e construções gramaticais; isto pode ser feito ao menos de duas maneiras: “Por registros: comparando entre diferentes tipos de documentos: língua falada, ficção, imprensa e registro acadêmico; Por período histórico: comparando diferentes séculos entre o XIII e o XX” (<http://www.corpusdelespanol.org/> - tradução nossa).

Esse *corpus* também permite fazer buscas baseadas na semântica. Por exemplo:

[...] é possível comparar e contrastar as colocações de duas palavras diferentes e determinar a diferença com respeito ao significado destas palavras. Também é possível averiguar a frequência e a distribuição de sinônimos cerca de 30.000 palavras e comparar sua frequência de aparição em diferentes registros assim como em distintos períodos históricos e utilizar estas listas de palavras como parte de outras buscas. Finalmente, é possível criar de forma simples listas próprias de palavras relacionadas semanticamente e utilizar tais listas como parte da busca (tradução nossa).

O CREA, *corpus* público e gratuito, “abrange um período que vai de 1975 até a atualidade, contém mais de 200 milhões de *tokens*⁵, dentre os quais 10% são transcrições da língua oral (50% provêm da Espanha e 50% dos países hispano-americanos)”.

3) Resultado e Análise dos Dados

3.1) Presença de heterossemânticos na Coleção *Enlaces Español para jóvenes brasileños*

Após analisar a Coleção *Enlaces* (3 volumes), foram encontrados apenas oito heterossemânticos em um exercício de vestibular no vol. 2 (p. 184), a saber: “*palco*”, “*zurdo*”, “*largo*”, “*brinco*”, “*comedor*”, “*berro*”, “*cena*” e “*taller*”.

3.2) Testagem de ocorrência dos heterossemânticos no CREA

A seguir apresentam-se as estatísticas obtidas no CREA com relação aos oito heterossemânticos da coleção *Enlaces*.

Foi usado o seguinte critério de busca: Cronológico: sem recorte - Meio: miscelânea - Geográfico: todos – Tema: todos.

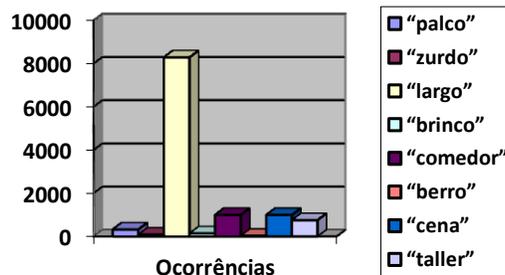
- a) “*palco*” - **Sem recorte cronológico:** 1 caso em 1 documento.
- b) “*zurdo*” - **Sem recorte cronológico:** 1 caso em 1 documento.

- c) “*largo*”- **Sem recorte cronológico:** 728 casos em 217 documentos.
- d) “*brinco*”- O CREA informa que não existem casos para esta consulta.
Sem recorte cronológico: 0 casos em 0 documentos.
- e) “*comedor*”- **Sem recorte cronológico:** 54 casos em 21 documentos.
- f) “*berro*”- O CREA informa que não existem casos para esta consulta.
Sem recorte cronológico temos: 0 casos em 0 documentos.
- g) “*cena*”- **Sem recorte cronológico:** 95 casos em 42 documentos.
- h) “*taller*”- **Sem recorte cronológico:** 69 casos em 39 documentos.

3.3) Testagem de ocorrência dos heterossemânticos no *Corpus del Español*

Vejam agora uma tabela e um gráfico com as estatísticas de cada heterossemântico da coleção *Enlaces* vol. 2:

Heterossemântico	Ocorrências
“ <i>palco</i> ”	323
“ <i>zurdo</i> ”	92
“ <i>largo</i> ”	8286
“ <i>brinco</i> ”	146
“ <i>comedor</i> ”	997
“ <i>berro</i> ”	29
“ <i>cena</i> ”	993
“ <i>taller</i> ”	756



3.4) CREA X CORPUS

Compilando e comparando as ocorrências dos heterossemânticos encontrados na Coleção *Enlaces* vol. 2 no CREA e no *Corpus del Español*, temos:

a) “*palco*”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
1	1	323

b) “zurdo”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
1	1	92

c) “largo”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
728	217	8286

d) “brinco”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
0	0	146

e) “comedor”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
22	12	997

f) “berro”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
0	0	29

g) “cena”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
95	42	993

h) “taller”

CREA		CORPUS DEL ESPAÑOL
CASOS sem recorte cronológico	DOCUMENTOS	OCORRÊNCIAS
69	39	756

Constatou-se que na coleção *Enlaces*, apenas o vol. 2 apresentava heterossemânticos, totalizando somente oito FA em todo o volume. Dos oito apresentados todos possuem uma única ocorrência, sendo que o vocábulo “brinco” foi a palavra que mais apresentou ocorrências em todo o volume, totalizando 68. Os demais heterossemânticos não possuem exemplos; porém, o mais curioso é que esse não foi o heterossemântico de maior ocorrência segundo as buscas feitas no CREA (sem recorte cronológico) e no *Corpus del*

Español. Esse vocábulo ficaria em 6º lugar de acordo com a ocorrência do *Corpus del Español* e em 6º lugar em ocorrência no CREA (vide dados abaixo). Dos vocábulos analisados, o heterossemântico que obteve mais ocorrência foi (ordem decrescente dando preferência aos dados do CREA):

- 1) “*largo*” totalizando 728 casos (CREA) e 8286 (*Corpus del Español*).
- 2) “*cena*” totalizando 95 casos (CREA) e 993 (*Corpus del Español*).
- 3) “*taller*” totalizando 69 casos (CREA) e 756 (*Corpus del Español*).
- 4) “*comedor*” totalizando 22 casos (CREA) e 997 (*Corpus del Español*).
- 5) “*zurdo*” totalizando 1 caso (CREA) e 92 (*Corpus del Español*).
- 6) “*palco*”, “*brinco*” e “*berro*” totalizando 1,0,0 casos (CREA) e 323, 146 e 29 respectivamente (*Corpus del Español*).

Após a compilação dos dados pode-se observar que nem sempre as ocorrências do *Corpus del Español* e do CREA se afinam. Vejam que o vocábulo “*taller*” ficou na 3ª colocação, porém no *Corpus del Español* ele fica em 4º lugar, enquanto que a palavra “*comedor*” de acordo com a ocorrência obtida pelo *Corpus del Español*, subiria de 4º para 2º lugar e o vocábulo “*cena*” desceria para o 3º lugar. Já o heterossemântico “*largo*” permaneceria na 1ª posição com ocorrência bastante alta em ambos os instrumentos de análise, enquanto que os heterossemânticos “*palco*”, “*brinco*” e “*berro*” aparecem empatados no CREA. Para uma melhor visualização da colocação dos FA, observe a tabela a seguir:

	CORPUS DEL ESPAÑOL	CREA sem recorte cronológico
1	“ <i>largo</i> ” (ocorrências= 8286)	“ <i>largo</i> ” (ocorrências= 728)
2	“ <i>comedor</i> ” (997)	“ <i>cena</i> ” (95)
3	“ <i>cena</i> ” (993)	“ <i>taller</i> ” (69)
4	“ <i>taller</i> ” (756)	“ <i>comedor</i> ” (54)
5	“ <i>palco</i> ” (323)	“ <i>zurdo</i> ” (1)
6	“ <i>brinco</i> ” (146)	“ <i>palco</i> ” (1) “ <i>brinco</i> ” e “ <i>berro</i> ” (0 sem recorte)
7	“ <i>zurdo</i> ” (92)	- X -
8	“ <i>berro</i> ” (29)	- X -

Assim sendo, pode-se observar que o heterossemântico com maior número de ocorrências na Coleção analisada, segundo as buscas efetuadas no *Corpus del Español* e no CREA, foi o vocábulo “*largo*”.

Considerações finais

Tendo em vista que um dos objetivos do EM é preparar os alunos para o Vestibular e para o ENEM e sabendo que os heterossemânticos fazem parte dos conteúdos das referidas provas, acredita-se que as autoras da coleção analisada deveriam explorar mais esse tema, o qual é considerado um dos mais difíceis no ensino/aprendizagem da Língua Espanhola. Igualmente, constatou-se que as autoras não se basearam nos números de ocorrências dos heterossemânticos ao fazer a seleção desses vocábulos. Além do que, dos inúmeros casos de heterossemânticos existentes no par de línguas espanhol/português brasileiro, foram usados tão somente oito exemplos em uma coleção de três livros. (Del 7 1)

Em função da importância deste tema no EM, Vestibular e ENEM, as demais etapas desta pesquisa têm como base as análises até aqui feitas para então trabalharmos: o questionário, a escolha de 72 ocorrências do heterossemântico “*largo*” no *Corpus del Español*, a tradução das 72 ocorrências no *Google Translate* e em seguida a retradução/ajuste, totalizando 144 ocorrências fazendo uso do dicionário escolar bilíngue impresso WMF. Lembrando que o vocábulo “*largo*” não foi escolhido aleatoriamente e sim por ele ter ficado em 1º lugar em ambas as ferramentas utilizadas na contagem de ocorrências: CREA (sem recorte cronológico) e *Corpus del Español*.

Dois motivos levaram à escolha das duas ferramentas de tradução acima mencionadas: 1º) O *Google Translate* estar à frente dos demais (*Bing*, *Translator da Microsoft*, *Babel Fish*, *Yahoo*); 2º) O dicionário WMF ser um dos instrumentos adotados por esta pesquisadora nas aulas de E/LE e o mesmo estar disponível na biblioteca do IFC – Campus Camboriú – SC. Mesmo presente durante as aulas, o dicionário não possui posição de destaque entre os recursos didáticos utilizados pelos docentes, por esse motivo, optou-se pelo uso de um dicionário impresso para fazer a retradução/ajustes nas traduções que serão feitas nas próximas etapas, por acreditar que ele é pouco, ou quase nunca utilizado em sala de aula.

Notas

¹ Palavras que levam o aprendiz a expressar-se de forma confusa e às vezes engraçada.

² Para o fim a que se destina este artigo, tomaremos como sinônimo os seguintes vocábulos: heterossemânticos, falsos cognatos e falsos amigos.

³ Todas as citações deste artigo retiradas de obras em língua estrangeira foram traduzidas por mim.

⁴ Justifica-se a escolha do *Google Translate*, tendo em vista o mesmo estar à frente dos rivais (*Bing*, *Translator da Microsoft*, *Babel Fish*, *Yahoo*) - <http://veja.abril.com.br/050510/lingua-google-p-122.shtml> acessado em 17/12/11 às 21:00 h - “O tradutor do *Google* está à frente dos rivais. Em pesquisas patrocinadas pelo governo americano, ele supera com frequência ferramentas de outras empresas e universidades”. “Ele também é reconhecido como o melhor entre os sistemas comerciais”, diz David Yarowsky, professor de ciência da computação da Universidade Johns Hopkins, nos Estados Unidos. Isso significa sobrepujar os rivais *Bing Translator*, da *Microsoft*, e *Babel Fish*, do *Yahoo*!

⁵ Em Linguística de *corpus*, as palavras “ocorrência” são chamadas de ‘tokens’ e os “vocábulos” de ‘types’ (SARDINHA, 2004, p. 165-166 *apud* MILLÁS, 2012, p. 51).

Referências bibliográficas

- ALMEIDA FILHO, José Carlos Paes de. **Linguística Aplicada - Ensino de Línguas e Comunicação** / José Carlos Paes de Almeida Filho - Campinas, SP: Pontes Editores e ArteLíngua, 3a Edição, 2009.
- ALONSO, M. C. G. P. - **Corpus Linguístico e a Aquisição de Falsos Cognatos em Espanhol como Língua** - Estrangeira, SP, 2006.
- ANDRADE, O. G. **Necessidades léxicas de universitários brasileiros aprendizes de espanhol: levantamento, descrição e análise**. 2010. 267 f. Tese (Doutorado em Estudos da Linguagem). Universidade Estadual de Londrina.
- BECHARA, Suely Fernandes – **¡Ojo con los falsos amigos!** – Dicionário de falsos cognatos em espanhol e português / Suely Fernandes Bechara, Walter Gustavo Moure; ilustrações Laert. – São Paulo: Moderna, 1998, p. 10.
- BERBER SARDINHA, A.P. - **Pesquisa em linguística de Corpus com WordSmith Tools** – Campinas, SP: Mercado de Letras, 2009.
- BOHN, I.& VANDRESEN, P. (org.). **Tópicos em linguística aplicada: o ensino de línguas estrangeiras**. Florianópolis: UFSC, 1988.
- BRASIL. Lei 9.394, de 20 de dezembro de 1996. **Estabelece as Diretrizes e Bases da Educação Nacional**. Diário Oficial da República Federativa do Brasil. Brasília, a. 134, n. 248, p. 27.836, 23 dez. de 1996.
- DURÃO, Adja Balbino de Amorim Barbieri- **Análisis de errores: en la interlengua de brasileños aprendices de español y de españoles aprendices de portugués**. 2 ed. Londrina: Eduel, 2004.
- FERNÁNDEZ, Francisco Moreno Jaime Otero Roth. **Atlas de la lengua española en el mundo**. Barcelona: Ariel -Fundación Telefónica, 2007. - Instituto Cervantes - El español: una lengua viva – Informe, 2012.
- FIALHO, V. R. e COSTA, J. S. - **O que o professor está fazendo? Análise sobre o uso de dicionários em sala de aula de línguas** - GARCIA, Maria de Los Angeles J., Josephine Sanchez Hernandez. **Español Sin Fronteras** – vol. 4 - Curso de Lengua Española para extranjeros. Ensino Fundamental - São Paulo: Scipione, 2007.
- JUNIOR, Pedro A. S. - **Análisis de errores. estudio de las estructuras verbales y discursivas en el aprendizaje del español – LE por parte de alumnos brasileños** - Salamanca, 2010.
- KREBS, Marloá Eggres – **Um olhar para a interferência dos heterossemânticos na aprendizagem do espanhol por falantes nativos do português brasileiro: um estudo de caso**. 146 f. Dissertação - Mestrado em Letras, Universidade Católica de Pelotas/RS, 2007.
- MENDES, Flávio Ramos. **Tecnologia e a construção do conhecimento na sociedade da Informação** - Londrina, 2007.
- MORAES, Irany Novah. **Elaboração da pesquisa científica**. Rio de Janeiro. Ed. de Publicações Médicas, 1978.

OSMAN, S. A. – **Coleção Enlaces: Español para jóvenes brasileños** - Soraia Adel Osman, Neide Elias, Sonia Izquierdo Merinero, Priscila Maria Reis e Jenny Valverde. Ed: Macmillan. Nível de ensino: Médio, 2010.

PROPOSTA CURRICULAR DE SANTA CATARINA. Disciplinas Curriculares Florianópolis. SED, 1998.

SILVANA, C.L. - **Dicionário escolar WMF: espanhol-português, português-espanhol** / [coordenação e revisão da tradução Silvana Cobucci Leite, Ivone C. Benedetti]. – São Paulo: Ed.: WMF Martins Fontes, 2011.

TAGNIN, S. E. O. **COMET: um Corpus Multilíngue para Ensino e Tradução**. São Paulo: USP, 2001, manuscrito.

_____. **Avanços da linguística de corpus no Brasil** – São Paulo: Humanitas, 2008.

Sites

ABRALE-

<http://www.abrale.com.br/biblioteca/O%20ensino%20do%20espanhol%20na%20educacao%20basica%20-%20desafios%20e%20perspectivas.pdf> - acessado em 03/12/2011 às 23:00h.

ALUÍSIO, S.M; Almeida, G.M.B. **O que é e como se constrói um corpus?** Lições aprendidas na compilação de vários corpora para pesquisa linguística - Disponível em: http://www.unisinos.br/publicacoes_cientificas/images/stories/pdfs_calidoscopio/vol4n3/art04_aluisio.pdf - acessado em 21/04/2013 às 14:00 h, p.156.

ANDRADE NETA, Nair Floresta. **Aprender espanhol es fácil porque hablo portugués: Ventajas y desventajas de los brasileños para aprender español**. Cuardeno Cervantes, Madrid - http://www.cuadernos cervantes.com/lc_portugues.html - acessado em 13/02/2012 às 23:07 h.

BERBER SARDINHA, A.P. - **Corpora eletrônicos na pesquisa em tradução** - http://www2.lael.pucsp.br/~tony/temp/publications/2003_Corpora_eletronicos_pesquisa_traducacao.pdf - acessado em 01/05/2013 às 13:57 h.

_____. **A influência do tamanho do corpus de referência na obtenção de palavras chave**. DIRECT Papers 38. São Paulo: PUC/SP. 1999a. Disponível em: <http://sites.uol.com.br/tony4/homepage.html> acessado em 24/06/2008.

_____. **Usando WordSmith Tools na investigação da linguagem**. DIRECT Papers 40. São Paulo: PUC/SP. 1999b. Disponível em: <http://sites.uol.com.br/tony4/homepage.html>; acessado em 24/06/2008 às 16:15 h.

_____. **Linguística de corpus**. São Paulo: Manole, 2004.

BRASIL, Ministério da Educação, Secretaria de Educação Média e Tecnológica. **Parâmetros curriculares nacionais, códigos e suas tecnologias. Língua estrangeira moderna**. Brasília: MEC, 1999. p. 49-63. Disponível em: <http://www.sk.com.br/pcn.html>. Acesso em: 13 nov. 2004.

CENTRO VIRTUAL CERVANTES DA ESPANHA. **Consejería de Educación y Ciencia. Mapa lingüístico de la lengua española en Brasil**. Brasília, Embajada de España, 1993. Disponível em: http://cvc.cervantes.es/obref/anuario/anuario_00/moreno/p03.htm acessado em: 26 maio 2012.

<http://www.cervantes.es/sobre_instituto_cervantes/publicaciones_espanol/espanol_mundo/anuario_2009.htm> - acessado em 03/12/2011 às 22:00 h (tradução nossa).

<<http://ultimosegundo.ig.com.br/mundo/instituto-cervantes-apresenta-anuario-e-brasil-e-destaque-no-ensino-do-idioma/n1237592106669.html>> - acessado em 03/12/2011 às 22:09 h.

<http://cvc.cervantes.es/lengua/anuario/anuario_00/moreno/p03.htm> acessado em 03/12/2011 às 21:53 h.

CORPUS DE REFERENCIA DEL ESPAÑOL ACTUAL (CREA) - Disponível em: <http://corpus.rae.es/creanet.html>.

CORPUS DEL ESPAÑOL. Disponível em: <http://www.corpusdelespanol.org/x.asp>.

DURÃO, Adja Balbino de Amorim Barbieri. **Por uma lexicografia bilíngue contrastiva** – Londrina: UEL, 2009. Disponível em:

<<http://www.europapress.es/portaltic/internet/noticia-espanol-segunda-lengua-materna-mundo-tercera-internet-20091102135432.html>> acessado em 03/12/2011 às 22:35 h – (tradução nossa).

GERBER, Regina Márcia; Vera Vasilévski (organizadora) – **Um percurso para pesquisas com base em corpus** – Florianópolis: Ed. Da UFSC, 2007, p. 179, 180.

INSTITUTO CERVANTES DA ESPANHA. Disponível em: http://www.cervantes.es/seg_nivel/biblioteca/Paises/Brasil/IBEROLENGUAS%20español%20en%20Brasil.mht - Acessado em: 25/05/12 às 10:00 h.

PARÂMETROS CURRICULARES NACIONAIS. Disponível em: <<http://www.mec.gov.br/sef/estrut2/pcn/pdf/livro01.pdf>> - acessado em: 10/10/04 às 19:00 h.

PGET, UFSC - Disponível em: http://www.pget.ufsc.br/curso/estrutura_curricular.php acessado em 07/07/2013 às 12:30 h.

SABINO, M. A. - **Falsos cognatos, falsos amigos ou cognatos enganosos?** Desfazendo a confusão teórica através da prática - Disponível em: <http://seer.fclar.unesp.br/index.php/alfa/article/view/1422> - acessado em 16/11/2012 às 15:00 h - Revista de Linguística, v. 50, n. 2, 2006.

SILVA, Eliane Barbosa da. **Bloqueios do aprendiz de espanhol/LE: os heterossemânticos**. In: CONGRESO BRASILEÑO DE HISPANISTAS, 2., 2002, San Pablo. **Proceedings online**. Associação Brasileira de Hispanistas, Available from: http://www.proceedings.scielo.br/scielo.php?script=sci_arttext&pid=MSC000000001200200100020&lng=en&nrm=abn acessado em 04/12/2011 às 11:33 h e 11/07/2012 às 15:00 h.

ZAPPAROLI, Z M - **A Linguística de Corpus e o programa WordSmith Tools** - Disponível em: <http://www.fflch.usp.br/dl/li/x/?p=193> - acessado em 01/05/2013 às 10:39 h.