

O léxico e seu uso em textos de alunos de língua estrangeira: um estudo descritivo pautado na Linguística de *Corpus* como subsídio para o ensino

(Lexicon and its use by foreign language students: a descriptive study
based on *Corpus* Linguistics to assist students with learning)

Celso Fernando Rocha¹

¹Faculdade de Ciências e Letras de São José do Rio Preto, UNESP - Universidade Estadual Paulista
celsotrad@yahoo.com.br

Abstract: The aim of this article is to present part of the results of a study concerning the lexicon employed by students learning Spanish language in a Languages degree (Undergraduate Education program). We describe the use and the context in which two verbal forms occur. We made use of *Corpus* Linguistics theory in order to compile two *corpora* of descriptive and argumentative compositions and observe the use of two Spanish verbs (*haber* and *tener* – third person singular). 250 compositions from first and second year students were collected. The *WordSmith Tools* software was applied to generate the list of words and the list of concordance. The verbal forms *hay* and *tiene* were the most used, and in some cases they were applied inappropriately when compared to the traditional Spanish grammar and to an electronic *corpus*. The results were discussed in class and were important to raise consciousness in relation to the students' textual production.

Keywords: Applied Linguistics; *Corpus* Linguistics; Spanish learners' *corpora*; lexicon and teaching.

Resumo: O objetivo deste artigo é apresentar os resultados parciais de um levantamento lexical de redações de aprendizes de língua espanhola de um curso de licenciatura em Letras. Procuramos descrever o uso e o contexto nos quais duas formas verbais ocorrem. Mais especificamente, recorreremos à Linguística de *Corpus* para compilar dois *corpora* compostos de textos descritivos e argumentativos e observar o emprego dos verbos *haber* e *tener*, conjugados na terceira pessoa do singular. Com relação aos passos metodológicos adotados, foram coletadas 250 redações de alunos do primeiro e segundo anos do referido curso. O programa *WordSmith Tools* foi utilizado para extração das listas de palavras e concordância. Verificou-se que as formas verbais *hay* e *tiene* foram as mais empregadas, e, em alguns casos, seus usos não corresponderam ao prescrito pela gramática tradicional espanhola e ao apresentando em um *corpus* de espanhol. Cabe mencionar que os resultados desse levantamento foram levados à sala de aula e serviram de subsídio para a reflexão sobre a produção textual dos discentes.

Palavras-chave: Linguística Aplicada; Linguística de *Corpus*; *corpus* de aprendizes de espanhol; léxico e ensino.

Introdução

Neste artigo, temos como escopo apresentar os procedimentos teórico-metodológicos e os dados parciais de um estudo maior de investigação do léxico em textos de aprendizes de espanhol como língua estrangeira (E/LE), obtidos por meio de um trabalho de coleta de 250 redações, escritas por dois grupos de alunos do curso de licenciatura em Letras com habilitação em língua espanhola (aproximadamente 15 alunos por grupo, falantes nativos de língua portuguesa). Objetivamos, mais especificamente, identificar as

duas formas verbais mais frequentes nos dois grupos e verificar os cotextos e contextos nos quais foram empregadas, contrastando-as com a norma padrão e com um *corpus* eletrônico de E/LE. Dessa forma, fazemos uso do instrumental disponibilizado pela Linguística de *Corpus* (LC) com intuito de, em um segundo momento, elevar o grau de conscientização do aprendiz, por meio da inserção em sala de aula dos dados extraídos dos *corpora*. Tal etapa de utilização dos dados para o ensino já começou a ser levada a cabo e, portanto, apresentaremos algumas reflexões iniciais sobre o procedimento adotado e seus resultados iniciais.

No que diz respeito à LC, pode-se dizer que vem estabelecendo relações com a Linguística Aplicada: ensino de língua estrangeira há algum tempo. Notamos, no ambiente pedagógico, a presença da LC em, pelo menos, quatro áreas de concentração, a saber: a) descrição de língua nativa, b) descrição de linguagem de aprendiz, c) ensino da metodologia de pesquisa em LC para alunos, e d) desenvolvimento de materiais didáticos.

Na primeira área, congregam-se o estudo da escrita e da linguagem oral de falantes nativos. Os desdobramentos desse tipo de pesquisa apresentam reflexos na produção de livros didáticos e de dicionários que são utilizados por professores e alunos. Desse modo, os levantamentos, análises e descrições linguísticas gerados pela LC estão indiretamente presentes em sala de aula por meio do material utilizado. O professor, nesse contexto, não faz uso direto da LC.

A segunda área de investigação possibilita diversos tipos de pesquisas, bem como de aplicação, em sala de aula, do conhecimento gerado. Para o docente, surge a possibilidade de coletar, armazenar em computador e descrever (qualitativa e quantitativamente) a linguagem produzida por aprendizes. Esses conjuntos de dados podem servir de referência para a produção de material de apoio, para o reforço de algum tópico desenvolvido em aula e para o planejamento de lições, com base nas dificuldades linguísticas mais frequentes dos alunos.

A terceira área estabelece relações com ensino explícito da metodologia da LC para alunos de língua estrangeira (LE). O instrumental analítico da LC auxilia os alunos (professor ou tradutor em formação) na aprendizagem/aquisição de léxico com o qual não tenha tido contato prévio; na compilação de *corpora* (comparáveis ou paralelos) de textos de determinadas áreas de especialidade; na elaboração de glossários bilíngues (espanhol e português); na preparação de aulas e de exercícios com base no léxico mais frequente, por exemplo.

A última área mencionada compreende a criação de abordagens ou metodologias de ensino que tomam como base conceitos advindos da LC. Na literatura da área, destacam-se, pelo menos, três modelos de aproveitamento do conhecimento gerado pela LC: o Currículo Lexical (WILLIS, 1990), a Abordagem Lexical (LEWIS, 1993) e o Ensino Movido por Dados (JOHNS, 1994).

A importância de pesquisas dessa natureza, levando em consideração os aspectos descritos anteriormente, reside no fato de que o mapeamento do conjunto lexical do *corpus* de aprendizes e a identificação de dificuldades mais comuns enfrentadas por graduandos de E/LE, ao empregarem o léxico e a gramática em LE nos seus textos, oferece subsídios para o direcionamento mais adequado do foco instrucional. Em outras palavras, aulas

com maior atenção dada às subáreas nas quais os aprendizes apresentem maiores dificuldades, levando-se em conta suas necessidades reais.

Corpus: algumas definições introdutórias

Para a LC, o termo “*corpus*” possui diversas definições: originalmente significava qualquer coleção de texto em meio digital ou não, organizado de acordo com determinados padrões (BAKER, 1995). A definição sofreu mudanças e passou a significar uma coleção de textos em formato eletrônico, passíveis de serem analisados automaticamente ou semiautomaticamente. Além disso, *corpus* deixou também de significar apenas textos de língua escrita para incluir textos orais e, por isso, é comum encontrarmos *corpora* tanto da fala quanto da escrita de iniciantes em línguas estrangeiras, por exemplo.

Atualmente, a presença de *corpora* torna possíveis análises que não teriam grandes possibilidades de serem executadas há algumas décadas, dadas as dificuldades advindas dos contextos teórico, metodológico e tecnológico. Na década de 1960, as palavras eram transferidas manualmente para cartões perfurados, para serem lidas por meios eletrônicos. Em tal momento, o estudo de determinadas estruturas linguísticas e posterior levantamento de frequências e contrastes apresentavam restrições de execução, devido à abrangência e à constituição do objeto de estudo.

Para Berber Sardinha (1999), uma definição mais abrangente, por incorporar as características principais para a construção de *corpora* em formato eletrônico, é a de Sánchez:

[...] um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SÁNCHEZ, 1995, p. 8-9 apud BERBER SARDINHA, 1999, p. 12)

Ainda, com relação ao *corpus*, sua natureza e seu tamanho permanecem variáveis, adaptando-se às situações e às perguntas de pesquisa, já que se podem ter *corpora* das mais diversas áreas e com variados propósitos. Cabe mencionar que sua compilação exige a observação de pré-requisitos operacionais, como: extensão, representatividade, especificidade e adequação.

Ao tratar do tamanho dos *corpora*, Berber Sardinha (2004, p. 26) argumenta que, apesar de ser um critério importante para sua compilação, há poucas pesquisas que abordam tal característica. O autor menciona três abordagens vigentes para determinar o número adequado de palavras em um *corpus*. A primeira, apresentada como “impressionística”, está baseada na observação e nas afirmações feitas por especialistas. A segunda, “histórica”, leva em consideração a monitoração dos *corpora* efetivamente utilizados pela comunidade ao longo de um período específico. A última tem como fundamento a “estatística”, para definir a quantidade necessária de palavras, a fim de retratar determinados aspectos ou características da língua.

A extensão e a representatividade estão interligadas. Entretanto, não dispomos de critérios objetivos para determinar o quão representativo é um *corpus*, pois, em última

instância, “o ônus de demonstrar a representatividade da amostra e de ser cuidadoso em relação à generalização dos seus achados para uma população inteira é dos usuários” (BERBER SARDINHA, 2004, p. 25).

A continuação, apresentaremos algumas conceituações em relação ao *corpus* de aprendizes.

Corpus de aprendizes e algumas pesquisas na área

Há muito tempo os estudos baseados em *corpora* influenciam o ensino de LE, em especial, de inglês. No século XX (até meados da década de 1980), seis nomes destacaram-se na aplicação de princípios da LC na produção de listas de frequências, feitas manualmente, para servirem de suporte à criação de material didático: Edward Thorndike, Herold Palmer, A. S. Hornby, Irving Lorge e Michael West (cf. BERBER SARDINHA, 2004, p. 252). Edward Thorndike, por exemplo, foi responsável pela publicação do *Teacher's workbook* (1921), levantamento compilado a partir de 4,5 milhões de palavras. O material propagou o ensino da língua materna e estrangeira com base no controle do vocabulário, ou seja, os alunos entrariam em contato com o léxico mais frequentemente empregado. Palmer e Hornby trabalharam na produção de uma *Interim Report on Vocabulary Selection*, que, posteriormente (1953), foi assumido por West (incorporando os trabalhos de Thorndike, Lorge, Palmer e Hornby). O pesquisador foi responsável pela *General Service of English Words* (duas mil palavras mais empregadas do inglês e comparadas com as definições do *Oxford English Dictionary*).

A empreitada levada a cabo pelos pesquisadores mencionados merece destaque, uma vez que a coleta e análise ocorreram em uma época em que não havia computadores. Mesmo na atualidade, coletar milhões de palavras e proceder a análises, com o auxílio do computador, não é tarefa simples.

De acordo com Granger (2002, p. 7), a compilação de *corpora* de aprendizes (*learner corpora* ou CCL, abreviatura para *Computer Learner Corpora*) é uma tarefa que exige muito esforço e tempo inestimável por parte dos pesquisadores: por isso, é importante ter um plano específico de coleta, armazenagem e extração de dados. Cabe salientar que os resultados advindos desse tipo de pesquisa são relevantes, pois podem auxiliar no desenvolvimento de materiais e na elaboração de currículos mais adequados às realidades culturais dos alunos. Segundo Berber Sardinha (2004), o desenvolvimento de materiais didáticos a partir da perspectiva de falantes nativos pode produzir uma visão idealizada da aprendizagem de LE, deixando de explorar as particularidades inerentes ao aprendizado de um grupo específico de alunos. O autor também menciona que, ao compilar *corpora* de aprendizes e explorá-lo por meio de ferramentas computacionais, o professor pode tomar decisões mais pontuais no que concerne à identificação e ao encaminhamento das dificuldades mais comuns enfrentadas pelos alunos.

Neste sentido, estudar as características da língua utilizada pelo aprendiz de LE sob a perspectiva da LC e por meio do *corpus* de aprendiz propicia um ponto de observação privilegiado. Segundo Lavid,

La disponibilidad actual de grandes colecciones de textos en formato digital y de herramientas informáticas ha hecho posible la realización de estudios sobre las regularidades con las

que los hablantes utilizan los recursos gramaticales de una lengua, investigando la distribución de frecuencias de diferentes construcciones, y las relaciones entre estructuras gramaticales y otros factores lingüísticos y contextuales, por un lado, y los factores que afectan a la selección entre las posibles variantes estructurales, por otro. (LAVID, 2005, p. 336)

Nos últimos anos, estudos que utilizam *corpus* passaram a disponibilizar descrições mais precisas de diferentes registros (linguagem oral, discurso formal, discurso acadêmico, diferentes dialetos, etc.). Estudos em língua inglesa demonstraram que existem padrões de uso de língua diferentes entre nativos e não-nativos. Petch-Tyson (1998), por exemplo, estudou a visibilidade do autor em quatro subcorpora (holandês, finlandês, francês e sueco), observando o emprego do pronome de primeira pessoa I e sua co-ocorrência com outros itens linguísticos. A autora descobriu que o uso de expressões como *I believe that*, *In my opinion* e *I am not of the same opinion* mostrou-se mais frequente em textos de não-nativos. Verificou-se, por conseguinte, uma transposição de expressões típicas da fala para o meio escrito pelos aprendizes de língua inglesa. Outros estudos também apontam diferenças do conjunto lexical utilizados por falantes não-nativos de inglês, dando especial atenção à interlíngua.

Por sua vez, em língua espanhola, podemos mencionar o trabalho pioneiro de Soto Balbás (2003), que faz uso de um arcabouço teórico-metodológico da LC para identificar, em um *corpus* de 25 mil palavras, os erros mais frequentes em 216 redações de aprendizes E/LE, nível básico. Também temos conhecimento de Cintrão (2009), que compilou um *corpus* de aprendizes E/LE, com foco na descrição de competência tradutória, e o projeto levado a cabo na Universidade Autônoma de Barcelona, intitulado *Corpus Escrito del Español L2* (CEDEL2), por meio do qual são descritos aspectos de aprendizagem de E/LE por falantes de língua inglesa.

Abordaremos, a continuação, os aspectos metodológicos adotados na confecção do estudo.

Passos empregados para a compilação dos *corpora* e extração dos dados

Com o intuito de analisar o uso dos verbos “haver” e “ter” (conjugados na terceira pessoa do presente do indicativo; uma vez que foram as formas mais frequentes na lista de palavras geradas pelo *WordSmith Tools* (doravante WST)) empregados por alunos de E/LE de um curso de licenciatura em Letras (primeiro e segundo anos – aproximadamente 15 alunos por turma) em 250 redações, procederemos à compilação de um *corpus* de aprendiz, composto por textos escritos no período de junho a dezembro de 2011. O quadro, a seguir, apresenta o número de redações, o prazo de entrega, os temas e o número de palavras por *corpus* coletado:

Quadro 1. Número, temas, tipos de redações coletadas e quantidade de palavras

Número da redação/ mês de entrega	Temas <i>Corpus I</i>	Temas <i>Corpus II</i>
01/ junho de 2011	Como eu sou	Como eu sou
02/ junho de 2011	Como meu melhor amigo é	Meio ambiente
03/ agosto de 2011	A organização pessoal	Descrição sobre a Internet
04/agosto de 2011	Descrever uma foto	Descrição sobre os sentidos
05/setembro de 2011	A casa dos sonhos	Enredo de novela
06/setembro de 2011	A vida em marte	Uma cidade
07/outubro de 2011	O zoológico	O que é a liberdade?
08/outubro de 2011	Uma cidade	Uma foto (feita em lab.)
09/novembro de 2011	O que é a liberdade?	Organização pessoal
10/novembro de 2011	Quais são suas lembranças?	
11/dezembro de 2011	Uma cidade (feita em lab.)	
Números	11 redações: 43.474 palavras	9 redações: 50.306 palavras

Quanto à extensão dos textos, pedimos para que os alunos escrevessem redações com no mínimo 300 e no máximo 500 palavras e as salvassem em formato .txt (textos sem formatação). Quando alguma redação não atingia o número mínimo, entrávamos em contato com o aluno e discutíamos as possibilidades de ampliá-la, ensinando-o a organizar suas ideias e a explorar melhor o tema. Por outro lado, quando o texto excedia o número máximo de palavras, era incorporado ao *corpus* sem nenhum corte. Dos 11 temas de redação do *corpus I*, um foi feito no laboratório de informática (uma cidade); do *corpus II* a redação que descrevia uma foto também foi escrita em laboratório, sem pré-levantamento de informações, consultas às obras de referências ou preparação de rascunhos. A intenção foi contribuir para o aumento da heterogeneidade na construção dos *corpora*. Pretendemos, futuramente, estudar a influência das condições de produção na qualidade do texto produzido.

Para as outras redações, o aluno recebeu o tema com um mês de antecedência e pôde realizar rascunhos em casa e enviar o texto por correio eletrônico. Foi permitido pesquisar em livros, dicionários e sítios da internet, bem como consultar o professor e os colegas. Houve vídeos e momentos de discussão sobre os temas em sala de aula, além de explicitação dos componentes essenciais de um texto descritivo ou argumentativo. Cabe salientar que a maioria dos textos são descritivos e que nosso *corpora* com textos argumentativos está sendo compilado.

A extração dos dados foi feita por meio do WST, um dos programas mais utilizados para pesquisas em LC. Esse *software*, criado por Michael Scott, professor da Universidade de Liverpool, possui três ferramentas: *WordList*, *Keywords* e *Concord*. No desenvolvimento deste trabalho, empregamos as ferramentas *WordList* e *Concord*. A ferramenta *WordList* permitiu criar listas de palavras por ordem de frequência e por ordem alfabética e, por isso, foi possível observar o léxico mais empregado no *corpus*.

Além das listas de palavras, a segunda ferramenta, a *Concord*, gerou listagens das ocorrências de itens específicos ou nódulos (verbos *haber* e *tener*, conjugados na terceira pessoa), acompanhados dos seus respectivos cotextos (texto ao redor da palavra), facilitando, assim, a identificação de sequências maiores nos textos (colocações). Essa ferramenta

também disponibilizou o utilitário *collocates*, responsável pela quantificação dos nódulos e identificação quantitativa de suas co-ocorrências. Dado o escopo da presente investigação, levantamos os primeiros dados referentes às coligações de *hay* e *tiene*.

Cabe acrescentar que as redações foram impressas, os “erros” (problemas de expressão escrita, inadequação lexical, regência verbal, concordância nominal, por exemplo) foram sublinhados manualmente, e os textos entregues aos alunos, sem maiores explicações. Caberia ao aluno, em sala, tentar corrigir seu texto. Caso não conseguisse, poderia perguntar ao colega e, posteriormente, ao professor ou ao grupo. Além dos trechos destacados nas redações impressas, levamos linhas de concordâncias geradas pelo WST para discussão em grupo. Desse modo, empregamos dois instrumentos para a extração dos dados: o WST e as discussões em sala. Também recorremos ao *corpus de referencia del español actual* (CREA, REAL ACADEMIA ESPAÑOLA, 2012), para constatar/ilustrar a possibilidade de alguns usos.

Apresentaremos, a seguir, os resultados obtidos por meio dos levantamentos efetuados.

Os verbos mais frequentes nos *corpora* analisados

Como mencionado anteriormente, apresentaremos os dados referentes às análises feitas a partir dos *corpora* selecionados. Para efetuar o estudo, as redações foram convertidas em arquivos .txt e inseridas no programa WST. A Tabela 1 apresenta esses verbos – em suas respectivas formas mais frequentes – nos textos dos *corpora* I e II.

Tabela 1. Verbos de maior ocorrência nos *corpora*

Verbo (forma mais frequente)	Ocorrência total	Corpus I	Corpus II
<i>Hay</i>	327	229	98
<i>Tiene</i>	286	187	99
<i>Era</i>	222	123	99
<i>Hacer</i>	207	89	118
<i>Tengo</i>	196	73	123
<i>Soy</i>	180	75	105

Por meio da Tabela 1, é possível observar o uso mais recorrente do verbo *haber* na terceira pessoa do singular do presente do indicativo (*hay* 229 ocorrências), no *corpus* I, e seu menor emprego (106) nas redações dos alunos do segundo ano de licenciatura em letras. O verbo *tener* (*tiene*) e o verbo *ser* (*era*), por sua vez, também na terceira pessoa do singular, registraram menor número de emprego no *corpus* II. Por outro lado, as formas *hacer*, *tengo* e *soy* foram mais utilizadas no *corpus* II, registrando, respectivamente, 118, 123 e 105 ocorrências. Observaremos, a seguir, os usos dos verbos *haber* e *tener* em nossos *corpora*.

Cálculo dos colocados mais frequentes nos *corpora* I e II com *hay*

Com o objetivo de analisar com maior acuidade as combinações que orbitam os verbos *haber* e *tener* (em suas formas conjugadas) nos dois *corpora*, procedemos ao le-

vantamento, por meio da ferramenta *concord* e de seu utilitário *collocates*, das construções prototípicas dos *corpora*. Em um primeiro momento, extraímos as colocações dos *corpora* e, em etapas futuras de pesquisa, interessa-nos (após a ampliação do *corpus*) estudar se há diferenças de emprego dessas combinatórias nas redações dos diferentes grupos de alunos.

Por ora, apresentamos a Figura 1, na qual se podem examinar os colocados mais frequentes no *corpus* I:

N	Word	With relation	Total	1st Left	2nd Left	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HAY	hay 0,000	245	0	0	4	1	3	0	0	229	0	0	3	1	4
2	DE	hay 0,000	79	36	44	3	1	22	0	0	0	1	0	12	0	13
3	Y	hay 0,000	73	22	51	4	4	7	3	4	0	0	5	22	13	11
4	LA	hay 0,000	67	49	18	10	4	6	29	0	0	2	1	1	5	0
5	UNA	hay 0,000	63	2	61	0	0	2	0	0	0	41	7	1	0	3
6	UN	hay 0,000	58	0	49	1	5	2	1	0	0	40	2	4	2	1
7	EN	hay 0,000	52	30	22	7	4	16	3	0	0	2	4	0	5	5
8	NO	hay 0,000	43	31	12	0	4	1	0	26	0	0	3	2	3	4
9	QUE	hay 0,000	41	19	20	2	1	1	3	6	0	1	4	14	6	3
10	EL	hay 0,000	39	24	15	4	5	4	11	0	0	4	3	3	4	1
11	CON	hay 0,000	27	5	22	0	3	2	0	0	0	0	0	0	11	3
12	DONDE	hay 0,000	26	14	12	1	1	0	4	0	0	0	1	4	6	1
13	TAMBIÉN	hay 0,000	24	6	18	0	0	0	0	6	0	16	0	1	0	1
14	MUCHOS	hay 0,000	16	0	16	0	0	0	0	0	0	15	0	0	0	1
15	SCORE	hay 0,000	16	14	2	1	1	0	3	0	0	0	0	1	1	0

Figura 1. Tela com o cálculo dos 15 primeiros colocados mais frequentes com *hay* no *corpus* I

O verbo *hay* apresenta-se com as seguintes combinações mais frequentes à direita: “[x] *no hay* [x]” (26 ocorrências), “[x] *donde hay* [x]” (8) e “[x] *también hay* [x]” (6). Já à esquerda, observa-se o maior número de colocados repetidos: “[x] *hay una* [x]” (41 ocorrências), “[x] *hay un* [x]” (40), “[x] *hay también* [x]” (15) e “[x] *hay muchos* [x]” (15).

Por sua vez, na Figura 2, podemos destacar o uso do mesmo verbo no *corpus* II:

N	Word	With relation	Total	1st Left	2nd Left	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	HAY	hay 0,000	100	1	1	0	1	0	0	0	91	0	0	0	1	0
2	QUE	hay 0,000	66	14	42	1	0	0	7	6	0	5	10	14	0	5
3	NO	hay 0,000	36	30	6	3	1	2	0	24	0	0	0	3	3	0
4	DE	hay 0,000	33	10	23	4	3	0	3	0	0	0	4	7	7	5
5	EN	hay 0,000	18	8	10	2	2	2	2	0	0	4	3	2	1	0
6	Y	hay 0,000	18	6	12	0	2	1	1	2	0	0	3	3	3	3
7	LA	hay 0,000	15	3	12	2	1	0	0	0	0	2	3	1	1	5
8	EL	hay 0,000	12	4	8	2	2	0	0	0	0	4	1	1	1	1
9	UN	hay 0,000	11	0	11	0	0	0	0	0	0	6	2	1	0	0
10	PERSONAS	hay 0,000	10	2	8	0	1	0	1	0	0	5	1	2	0	0
11	LO	hay 0,000	10	3	7	1	1	0	1	0	0	1	2	0	4	0
12	UNA	hay 0,000	9	1	8	0	0	1	0	0	0	6	2	0	0	1
13	LOS	hay 0,000	8	1	7	0	1	0	0	0	0	0	0	1	4	2
14	PERO	hay 0,000	8	7	1	0	0	0	0	7	0	0	0	0	1	0
15	POR	hay 0,000	7	1	6	0	0	0	1	0	0	1	1	3	0	1

Figura 2. Tela com o cálculo dos 15 primeiros colocados mais frequentes com *hay* no *corpus* II

De modo semelhante, a combinação “[x] *no hay* [x]” (24 ocorrências) foi a mais frequente, seguida de “[x] *pero hay* [x]” (7). À direita, observa-se o uso de “[x] *hay un* [x]” (8) e “[x] *hay una* [x]” (5).

Um aspecto importante a ser observado é o número de uso absoluto dos itens que orbitam a forma *hay*, apresentado na quinta coluna das figuras 1 e 2. Pode-se constatar

que, por exemplo, há emprego maior no *corpus* II de combinações que orbitam o relativo *que* e a partícula negativa *no*. Por outro lado, tais combinações ocorrem em frequência menor no *corpus* I, uma vez que as construções mais recorrentes são aquelas em que aparecem os artigos indeterminados. De modo semelhante, a forma *tiene* (ver figuras 3 e 4) é usada com maior frequência no *corpus* II em perífrases que expressam obrigações e com maior frequência no *corpus* I para expressar posse.

O fator seriação das turmas e a existência de alguns textos diferentes nos *corpora* podem explicar o uso apresentado e, também, indicar ao professor traços de sobreuso ou subuso de determinados padrões, além de contribuir para o desenvolvimento e a inserção de atividades específicas, atentando, por conseguinte, para as necessidades de aprendizagem em determinado grupo.

Cálculo dos colocados mais frequentes nos *corpora* I e II com *tiene*

A Figura 3, abaixo, ilustra as combinações mais usuais no *corpus* composto por redações dos alunos do primeiro ano:

N	Word	Verb	Total	Left	Right	L1	L2	L3	L4	L5	Center	R1	R2	R3	R4	R5
1	TU	tiene	109	1	1	0	1	0	0	0	107	0	0	0	1	0
2	Y	tiene	76	23	54	4	2	4	0	0	0	2	3	14	21	14
3	LA	tiene	54	33	21	0	3	0	0	0	0	0	0	2	3	0
4	DE	tiene	53	23	30	7	3	11	2	0	0	0	3	14	7	0
5	QUE	tiene	50	31	19	0	3	4	0	0	0	7	0	4	2	0
6	UNA	tiene	50	9	41	1	2	0	0	0	0	0	0	0	2	2
7	UN	tiene	36	6	30	2	4	0	0	0	0	0	0	0	1	4
8	EL	tiene	36	13	23	1	3	4	0	0	0	0	0	0	0	1
9	ES	tiene	34	23	11	7	0	0	0	0	0	0	1	0	1	4
10	LOS	tiene	22	5	17	1	0	2	2	0	0	0	0	1	1	0
11	EN	tiene	19	0	19	1	3	3	1	0	0	2	0	0	3	1
12	TAMBIÉN	tiene	17	0	17	0	0	0	1	0	0	0	0	0	0	1
13	NO	tiene	17	14	3	0	2	1	0	0	0	0	1	0	0	2
14	MI	tiene	16	16	0	4	0	2	0	0	0	0	0	0	0	0
15	POR	tiene	16	9	7	2	3	1	3	0	0	0	0	3	0	4

Figura 3. Tela com o cálculo dos 15 primeiros colocados mais frequentes com *tiene* no *corpus* I

Por meio da Figura 1, constatamos que as combinações mais frequentes no *corpus* foram: “[x] *tiene una* [x]” (30 ocorrências), “[x] *tiene un* [x]” (15), “[x] *tiene el* [x]” (14), “[x] *tiene la* [x]” (10), “[x] *tiene los* [x]” (7); no caso de *también* e *que*, verificam-se combinatorias à esquerda e à direita de forma alternada: “[x] *también tiene* [x]”/ “[x] *tiene también* [x]” (7+6) e “[x] *que tiene* [x]”/ “[x] *tiene que*[x]” (13+7).

No *corpus* II, chama-nos a atenção o uso pouco frequente de perífrases que expressam negação ou obrigação. Podemos constatar tal fato por meio da observação do resultado do cálculo de colocados na Figura 4:

N	Word	With	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	TIENE	tiene	99	0	0	0	0	0	0	0	99	0	0	0	0	0
2	QUE	tiene	47	32	15	1	4	10	6	11	0	3	0	2	5	5
3	DE	tiene	46	16	30	6	4	4	2	0	0	1	4	13	4	8
4	LA	tiene	40	19	21	2	3	5	9	0	0	12	0	1	3	5
5	Y	tiene	31	10	21	2	2	0	3	3	0	0	2	7	9	3
6	EL	tiene	29	14	15	3	2	5	4	0	0	7	3	1	4	0
7	EN	tiene	15	6	9	1	2	3	0	0	0	1	2	0	3	3
8	ES	tiene	15	11	4	1	5	1	4	0	0	0	0	0	2	2
9	NO	tiene	14	13	1	0	0	1	0	12	0	0	0	0	0	1
10	LOS	tiene	14	3	11	1	1	0	1	0	0	2	1	1	4	3
11	UN	tiene	13	6	7	2	1	1	1	1	0	5	1	0	1	0
12	UNA	tiene	12	3	9	1	0	1	0	1	0	6	1	0	0	2
13	LAS	tiene	10	3	7	3	0	0	0	0	0	3	0	1	1	2
14	MUY	tiene	9	4	5	1	0	2	1	0	0	0	1	4	0	0
15	PERO	tiene	9	6	3	0	2	1	1	2	0	0	0	1	2	0

Figura 4. Tela com o cálculo dos 15 primeiros colocados mais frequentes com *tiene* no *corpus* II

No que concerne às combinações com o uso de “que”, observa-se o emprego desse pronome relativo à direita em pelo menos 11 situações ([x] *que tiene* [x]) e 3 vezes com ele ou à direita e à esquerda ou só à direita ([*que tiene que* [...]/ *tiene que* [x]). A construção *tiene que* foi empregada com *decir*, *tener* e *ver* (no infinitivo). Outra construção mais frequente foi a que traz à direita o artigo feminino *la* ([x] *tiene la* + *substantivo feminino*) (*piel*, *capacidad*, *libertad*, *percepción*, etc.)

O cálculo das colocações demonstra que é frequente a combinação *no* + *tener* + *substantivo* (*miedo*, *tiempo*, *sentido*, etc.). Uma futura investigação poderia descrever as características de tais combinações, uma vez que, em alguns dos exemplos do *corpus*, parece haver emprego de *tener* como “coringa”; mais especificamente, por ser verbo de significação mais ampla, estaria sendo usado no lugar de outros mais específicos. Em um exercício realizado como *warm up*, escrevemos na lousa a frase: *Las personas no tienen tiempo para [...]* (exemplo retirado do *corpus* II) e pedimos para que os alunos substituíssem *tiene* por outro verbo e refletissem sobre a mudança. A escolha recaiu sobre o verbo *disponer* (*de*), aumentando, dessa forma, o grau de especificidade. Uma aluna mencionou que a frase “ficou mais espanhola e objetiva”, ou seja, notou que a combinação agrega maior precisão ao período e está mais próxima de um registro formal. Tais exercícios, além de promoverem reflexão, contribuem para o enriquecimento do repertório linguístico em LE.

Na próxima seção, encontram-se observações mais detalhadas sobre o emprego das formas verbais selecionadas.

Os usos de *hay* nas redações do *corpus* I e II

O verbo *haber*, em língua espanhola, apresenta-se com alguns usos peculiares, e seu emprego adequado pode trazer certa dificuldade para o aprendiz. Trata-se de um verbo irregular com uma forma especial na terceira pessoa do singular (*hay*); é utilizado, obrigatoriamente, como auxiliar na formação dos tempos compostos (*he sabido*, *has entendido*, etc.); em perífrases verbais (p.ex.: as de obrigação: *haber* + *de* + infinitivo); quando é importante mencionar a presença de ser(es) ou objeto(s) designados por um substantivo utilizado após *haber* (p. ex.: *Hay un cuaderno en la mesa.*), de modo a enfatizar a impessoalidade ou matizar características de um objeto qualquer.

Levando em consideração os usos mencionados, passamos a ler, por meio da ferramenta *Concord*, as 327 linhas de concordância nas quais a forma *hay* aparece. O

critério de seleção das linhas foi quantitativo (por meio do levantamento apresentado nas figuras 1 a 4 e da observação dos *clusters* mais frequentes) e qualitativo. Após a leitura, destacamos dez linhas com base em um número abrangente de casos que contemplassem nossa intuição, anotações durante as correções das redações e discussões em sala. A título de ilustração, seguem as dez primeiras linhas eleitas do *corpus* I:

Quadro 2. Linhas de concordância geradas do *corpus* I

Linha	Concordância
1	<i>En los dos lados de la cama hay mesillas de noches con una lámpara. Delante de la cama hay un escritorio y una silla. Sobre el escritorio hay mi ordenador, mis hojas, [...]</i>
2	<i>Hay un ropero blanco empotrado delante de la cama.</i>
3	<i>[...] cama doble con una sábana blanca con dibujos de mariposas violetas. Sobre la cama hay algunos cojines también blancos y con mariposas violetas.</i>
4	<i>Hay cuatro ventanas y una cortina blanca.</i>
5	<i>Algunas personas acreditan [sic] que ya hay vida en Marte. No sé, pero pienso que, se [sic] realmente hay, ellos están ciertos [...]</i>
6	<i>Hay un libro en la mesa de la cocina.</i>
7	<i>[...] agradables: cuidan del solo, de las plantas, viven una vida muy tranquila. No hay guerras, no hay problemas, los marcianos no erran [sic], siempre procuran reflejar</i>
8	<i>Al lado de la ventana hay un escritorio y una silla y sobre el escritorio hay un portalápices negro y también una lámpara azul.</i>
9	<i>Sin embargo, en el lado derecho de la casa está la sala de estar [sic], (hay) tiene otra sala también [...]</i>
10	<i>Rio Preto es muy grande, hay muchos coches, gentes y rascacielos.</i>

Cabe salientar que observamos as linhas e identificamos, quase na totalidade dos casos, o emprego de *hay* para mencionar objetos (ex. linhas 1 a 9) e, em outros, para aludir à existência de pessoas, atribuindo à oração um matiz de impessoalidade (ex. linha 10, *mucha gente* é complemento direto do verbo *haber*).

Na linha 1, por seu turno, verifica-se o uso da sequencia *hay mi ordenador, mis hojas [...]*, na qual o verbo *haber* foi empregado equivocadamente no lugar do verbo *estar*, pois o sintagma *mi ordenador* e *mis hojas* seriam sujeitos do verbo *estar*. O uso do pronome *mi* atribui um traço de singularidade ao objeto, impossibilitando o uso de *hay* (impessoal).

Na linha 5, há o emprego adequado do verbo *haber* e a utilização de *acreditar*, não apropriado ao contexto, pois, em espanhol, esse verbo tem o sentido de demonstrar o valor ou a autenticidade de certo objeto. Poderia ser traduzido para o português por “acreditar” (no sentido de dar crédito), “atestar” ou “assegurar”. O verbo adequado ao contexto seria *creer* (*yo creo que...*).

Na linha 7, verificamos o uso do verbo *erran*, que em língua espanhola, assim como em língua portuguesa, pode ter, pelo menos, dois sentidos: o primeiro é andar sem destino fixo ou objetivo (“vagar”) e o segundo é “desacertar”. Apesar de os sentidos terem sido registrados em dicionário, e o aluno ter utilizado uma obra de referência para argumentar que estava certo no emprego do verbo, sabíamos, por conhecimento intuitivo da língua, que o seu uso não estava adequado, ou, no mínimo, a forma não é frequente. Em buscas efetuadas no CREA, a forma conjugada *erran* registrou uma ocorrência com o sentido de

perambular sem destino ([...] *los seres de carne y hueso que, como una multitud de sombras, erran por sus novelas..* Por sua vez, a mesma forma verbal no infinitivo retornou 182 ocorrências com o sentido de “equivocar-se” (sentido constatado nas primeiras 25 primeiras linhas de concordância do CREA).

No trecho 9, há dúvida no uso de *tener* ou *haber*. Em conversas posteriores, o aluno relatou sua dificuldade em compreender e diferenciar o emprego dos dois verbos. Alguns exercícios foram oferecidos ao aluno, e, em outras redações, constatamos o emprego adequado dos dois verbos.

O quadro, a seguir, ilustra alguns usos mais frequentes do verbo *haber* pelos alunos do segundo ano de licenciatura em Letras:

Quadro 3. Linhas de concordância geradas do corpus II

Linha	Concordância
1	<i>Pero no es mi habitación. La habitación, aquella en que vivo, es diferente: en ella hay muchos colores. Hay el verde de mis ojos, que hoy no se puede ver, ya que [...]</i>
2	<i>Lo que devemos [sic] meter en la cabeza es que hay la parte buena y la mala en todo en la vida.</i>
3	<i>Hay seres humanos que, a través de las descubiertas científicas, hacen bombas para matar e [sic] destruir, en cuanto otros buscan la cura de enfermedades.</i>
4	<i>El paladar es muy personal, pues hay personas que no les gustan de salado [sic], hay otras que no les gustan de ácido [sic] tampoco y hay gente que les gusta de todo [sic]</i>
5	<i>Hay una alianza entre el hombre y la naturaleza, todo está interconectado</i>
6	<i>Pero hay el silencio, invisible, pero presente en la imagen.</i>
7	<i>Hay cines y centros comerciales.</i>
8	<i>Desafortunadamente, ya hay embotellamiento en mí ciudad porque esa ciudad se ha desarrollado mucho en lo últimos años.</i>
9	<i>La verdad es que no hay colores, ningún... No hay el amarillo de las falsas sonrisas, ni hay el oscuro de los ojos de los muertos. No hay, también, el rojo de las bocas de las putas. Hay sólo yo. Nadie más vive aquí.</i>
10	<i>A pesar de ser una persona extrovertida, hay veces que soy más reservada.</i>

No *corpus* II, podemos destacar alguns usos que estabelecem relações com contextos mais abstratos. Na linha 1, do Quadro 3, há o excerto de uma redação na qual o aluno descreve parte de sua casa. Verificamos o emprego do sintagma *hay muchos colores*, de forma adequada, remetendo não ao contexto físico da casa, e sim, a um estado de espírito ou o possível estado interior do aluno-autor. De modo análogo, na 6, há o uso de *hay el silencio*, sintagma utilizado em uma redação na qual o aluno descrevia uma foto com um “tom” tétrico. Na construção, sobressai o sentido abstrato e indeterminado do sintagma *el silencio*, apesar de estar sendo empregado um artigo determinado o restante da frase apresenta a contradição no plano discursivo, por meio das palavras *invisible* e *presente*. Criando, portanto, um jogo expressivo duplo, entre o conhecido e o desconhecido, entre o determinado e o indeterminado. Segundo Gili (2005, p. 159), a estrutura gramatical de uso seria: *verbo na terceira pessoa + artigo indeterminado ou um adjetivo indefinido e um substantivo*. No caso citado, há inserção de componente determinado gramatical (*hay el silencio*) e dois elementos lexicais, um que indetermina (*invisible*) e outro que acrescenta certo matiz de determinação (*presente*) ao tipo de silêncio (o uso de *pero* contribui

para o aumento do grau de indefinição). No CREA, foram encontrados dois usos para o sintagma *hay el silencio* (usos metafóricos) e mais de 150 usos para construção *hay silencio* (sem intercalação de artigo e em contextos metafóricos – *hay el silencio interno de las aves / de las leyes, de una habitación*). Cabe dizer que, de modo semelhante, o exemplo da linha 9 também explora os níveis de determinação e indeterminação (ou conhecido e não conhecido).

No exemplo da linha 2, por seu turno, o emprego do artigo determinado (... *hay la parte buena...*) não está adequado, e deveria ser substituído pelo artigo indefinido *una*, para que possa haver um nível de indeterminação e não especificação da “parte” da vida que é “boa” ou “má”. Seria difícil precisar quais seriam as “partes boas” da vida ou as “partes más”. Dessa forma, somente com o uso do indefinido a generalização dicotômica terá seu sentido expresso de forma “eficiente”. Nas linhas 3, 4, 7, 9 e 10, verifica-se o uso adequado de *hay* + substantivo plural. O verbo se mantém na terceira pessoa do singular.

Os usos de *tiene* nas redações do *corpus* I e II

Como consta da Tabela 1, o verbo *tener*, em sua forma conjugada de terceira pessoa do singular, registrou 286 ocorrências totais (187 + 99) nos *corpora* do presente estudo. Trata-se de um verbo irregular e intransitivo com o sentido de estabelecer relação entre uma pessoa e um objeto possuído. É utilizado em inúmeros contextos em construções como: *tener a alguien de ayudante, tenerlo como ayudante, tener a alguien por idiota, tenerla tomada con alguien; tener que + infinitivo*, para demonstrar necessidade, obrigação ou determinação para fazer aquilo que está expresso por meio do infinitivo. Na construção *tener de + infinitivo*, expressa vontade, e, em *tener + participio*, ressalta ato terminativo.

Por meio da ferramenta *Concord*, foram extraídas as linhas de concordância para a forma *tiene*. O quadro 4, na continuação, apresenta algumas linhas nas quais o verbo foi utilizado no *corpus* I.

Quadro 4. Linhas de concordância geradas do *corpus* I

Linha	Concordância
1	<i>El colegio tiene y tendrá una grande [sic] parte en mi memoria gastada de mi cabeza.</i>
2	<i>Apesar [sic] de pequeña, tiene casi todo lo que los habitantes necesitan para vivir:[...]</i>
3	<i>Ultimamente [sic] tiene ganado [sic] premios por su participación activa para [sic] la preservación del medio ambiente</i>
4	<i>Mariana tiene pelo liso, corto y negro, también tiene gafas y es baja y delgada.</i>
5	<i>Nuestro amigo se llama Andrés, tiene veintiocho años, es ingeniero [sic] estudia inglés. Él vive sólo en Córdoba y no tiene ninguna mascota. Su pelo es negro</i>
6	<i>Él tiene un perro.</i>
7	<i>Claro que tiene las personas malas, pero es una ciudad pequeña, con solo 150 mil habitantes [...]</i>
8	<i>En mi habitación tiene una cama [...]</i>
9	<i>De el [sic]comedor se tiene acceso a la cocina.</i>
10	<i>Para nosotros ese lugar tiene que ser preservado de la manera que Dios lo hizo.</i>

O uso mais frequente foi no sentido de atribuir a posse de um objeto/característica/animal a um ser animado. Também registramos uso de lugares (casa e cidade) como detentores de características ou objetos (exemplos 2 e 10).

Com relação aos usos menos frequentes, podemos mencionar o primeiro exemplo, no qual o verbo está atribuindo, de forma metafórica, a posse de parte da memória ao sujeito *colegio*. Em vez de o sujeito ser representado por uma pessoa detentora de memória e, portanto, capaz de registrar os fatos independentemente, o sujeito *colegio* passa a ser dono da memória do aluno-autor.

Na linha 3, o emprego de *últimamente* pressupõe o uso da seguinte construção: *estar* + gerúndio (ou reescrevendo com *pretérito perfecto*: *En este año la ciudad ha ganado muchos premios...*). Quando a expressão é utilizada como equivalente de *haber* + participípio, pode-se constatar um caso de regionalismo galego e asturiano (cf. SECO, 1989, p. 358).

Nas linhas 7 e 8, há o uso de *tiene* no lugar de *hay* (ou *existir*). Talvez por influência da língua portuguesa, uma vez que é frequente empregarmos “ter” no lugar de “haver” em registros menos formais (exemplo: “tem um carro na garagem” / “há um carro na garagem”). E, em outros contextos em língua portuguesa, podemos permutar uma forma pela outra, em casos em que “ter” e “haver” funcionem como verbos auxiliares (eles tinham chegado / eles haviam chegado).

É importante salientar que esperávamos maior ocorrência desse fenômeno, já que, em linguagem oral, durante as atividades de conversação em sala de aula, houve emprego acentuado de *tener* no lugar de *haber*; no entanto, após a leitura das linhas de concordância, elencamos apenas 3 ocorrências nas quais essa substituição foi feita.

Outros usos observados foram os das linhas 9 (menos frequente) e 10. No primeiro caso, há o uso de “se” apassivador, e, na última, a perífrase verbal *tener que*, expressando obrigatoriedade. O último ocorreu 7 vezes no *corpus* I.

No quadro abaixo, apresentamos os exemplos de uso retirados do *corpus* II:

Quadro 5. Linhas de concordância geradas do *corpus* II

Linha	Concordância
1	<i>Los colores de la naturaleza tan radiantes e iluminados por los rayos del sol parece [sic] nos dar las sugerencias de que nuestra vida tiene que tener muchos colores.</i>
2	<i>Ella tiene 19 años y vive en Catanduva y todos los días viaja para Rio Preto.</i>
3	<i>Tiene su cabello negro y lleno, que siempre lleva una cola de caballo.</i>
4	<i>Cada especie tiene el suyo lugar, cada una tiene la suya responsabilidad en la vida.</i>
5	<i>Sin embargo, sabemos que todo lo que nos rodea tiene sus lados positivos y negativos.</i>
6	<i>Tiene pelo rubio, lleva gafas, es un poco baja, pero no es gorda. Es una chica muy guapa.</i>
7	<i>La vida, ese gran ordenador, que tiene mucha memoria temporaria, pero poca memoria física.</i>
8	<i>Pero una cosa más llena esas calles, algo que no puedo explicar, pues es algo que no es visible, es algo que tiene que ver con el legado que esas pequeñas y estrechas sendas cargan.</i>
9	<i>Bueno, después te cuento más cosas, es que ahora Antonio y yo vamos a conocer otro sitio de la ciudad, parece que es otra finca de su familia, donde tiene varios animalitos y todo más; y él quiere mostrarme eso.</i>
10	<i>El internet es un servicio óptimo y, con atención, solo tiene a ayudar.</i>

Tanto no *corpus* I quanto no *corpus* II os usos mais frequentes estão relacionados à atribuição de posse de um objeto ou característica a um sujeito (linhas 2,3,4,5,6 e 7). E os usos menos frequentes dizem respeito ao emprego de *tener* em perífrase verbais (linha 1).

No exemplo da linha 3, a construção com o uso do adjetivo possessivo *su* acrescenta um traço de cuidado em relação à manutenção do cabelo. Em segundo plano, teríamos a constatação das propriedades do cabelo, ou seja, o fato de seu cabelo *ser* negro (*tiene el cabello negro* expressa, em espanhol, o mesmo sentido em português de “seu cabelo é negro”). Na linha 4, há redundância de informação por meio do emprego (equivocado) da forma tônica do pronome no lugar da forma átona. Como se sabe, a forma tônica é responsável pela inserção no contexto de informações novas a respeito de determinado substantivo. O emprego correto seria *cada especie tiene su lugar*, com a forma átona, introduzindo, por conseguinte, elementos não conhecidos.

Com relação aos usos menos frequentes, podemos mencionar o exemplo da linha 8, que traz uma construção, ressaltando o estabelecimento de relações entre um objeto e outro. A expressão também é utilizada na forma negativa (*no tener algo nada que ver con otro*).

No exemplo 9, por seu turno, registramos o único caso de emprego de *tener* no lugar de *haber*.

Observações finais

Ao empregarmos o arcabouço teórico-metodológico citado e compilarmos dois *corpora* de aprendizes de E/LE, pudemos acercar-nos de fenômenos colocacionais mais frequentes nos *corpora* analisados, além de iniciar a utilização dos dados levantados em nossas atividades didáticas, uma vez que alunos e docente iniciaram a observação das linhas de concordâncias selecionadas e passaram a refletir sobre o emprego dos dois verbos nas redações.

Em um dos momentos de reflexão, pudemos promover uma ampliação sobre a compreensão de aspectos relacionados ao uso do E/LE. Mais detalhadamente, a turma do primeiro ano começou a explorar as várias possibilidades de significação e as combinações de uso do verbo *haber*. Notaram que havia uma prescrição mais fechada, no material didático adotado, e um uso mais flexível e mais metafórico nos exemplos do CREA e nas próprias redações. A “descoberta”, de certo modo, surpreendeu vários alunos, pois alguns ainda acreditam (ou acreditavam) que as obras de referência trouxessem *todas* as formas de uso e normatizassem o uso do idioma estudado de maneira extensiva. Os contatos com a diversidade e a observação sistematizada do emprego inadequado do léxico passaram a despertar nos discentes uma inquietação salutar, ampliando, por conseguinte, o leque de consulta aos *corpora* eletrônicos e à *web*, além, claro, de manter as consultas aos dicionários e às gramáticas. Assim, consideramos que a primeira experiência de levar os dados gerados para serem discutidos em aula foi útil e será ampliada.

Em etapas posteriores da pesquisa – além de abordarmos a descrição de outros itens lexicais – trataremos das dificuldades no gerenciamento adequado de um *corpus* de aprendiz e de suas idiossincrasias; a importância da devolução em menor tempo das

redações aos alunos, do *feedback* consistente e do tratamento, em sala de aula, das dificuldades encontradas. Também merecerá espaço a discussão a respeito da forma de coleta, das instruções de organização da escrita para o aprendiz, entre outros aspectos.

Por fim, neste artigo, tivemos como objetivo apresentar os dados parciais de um estudo que se encontra em desenvolvimento e que tem como foco a descrição e o encaminhamento em sala de aula de questões interligadas à aprendizagem de E/LE.

Esperamos que a metodologia adotada possa ser empregada por outros pesquisadores e que os resultados possam servir de base para a reflexão sobre a aprendizagem de LE.

REFERÊNCIAS

BAKER, M. *Corpora* in Translation Studies: an Overview and Some Suggestions for Future Research. *Target*, Amsterdam, v. 7, n. 2, p. 223-243, 1995.

BERBER SARDINHA, A. P. A influência do tamanho do *corpus* de referência na obtenção de palavras-chave. *DIRECT Papers*, 38; São Paulo: PUC/SP, 1999. Disponível em: <<http://sites.uol.com.br/tony4/homepage.html>>. Acesso em: 24 jun. 2005.

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

CINTRÃO, H. Um *corpus* de tradução no par português-espanhol. *Tradução e Comunicação*, São Paulo, n. 18, p. 131-144, 2009.

GILI, O. C. *Diccionario práctico de gramática*. Madrid: Edelsa, 2005.

GRANGER, S. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 2002.

JOHNS, T. From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. In: ODLIN, T. (Ed.). *Perspectives on Pedagogical Grammar*. New York: Cambridge University Press, 1994. p. 293-313.

LAVI, J. *Lenguaje y nuevas tecnologías*. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI. Madrid: Cátedra, 2005.

LEWIS, M. *The lexical approach: the state of ELT and a way forward*. Hove: Language Teaching Publications, 1993.

PETCH-TYSON, S. Writer/reader visibility in EFL written discourse. In: GRANGER, S. (Org.). *Learner English on Computer*. Nova York: Longman, 1998. p. 107-118.

REAL ACADEMIA ESPAÑOLA. *Corpus de Referencia del Español Actual (CREA)*. 2012. Disponível em: <<http://corpus.rae.es/creanet.html>>. Acesso em: 8 jul. 2012.

SÁNCHEZ, A. Definición e historia de los *corpus*. In: SÁNCHEZ, A. et al. (Org.) *Cumbre: Corpus lingüístico del Español contemporáneo*. Madrid: SGEL, 1995. p. 7-24.

- SCOTT, M. *WordSmith Tools*. Version 4. Oxford: Oxford University Press, 2006.
- SECO, M. *Diccionario de dudas y dificultades de la lengua española*. Madrid: Espasa Calpe, 1989.
- SOTO BALBÁS, M. *Análise de erros, baseada na Lingüística de Corpus, da escrita de aprendizes brasileiros universitários de Espanhol como Língua Estrangeira*. 2003. 128f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica de São Paulo, PUC-SP, São Paulo.
- THORNDIKE, E. L. *Teacher's wordbook*. Nova York: Columbia Teachers College, 1921.
- WEST, M. *A general service list of English words*. Londres: Longman, 1953.
- WILLIS, D. *The lexical syllabus: a new approach to language teaching*. London/Glasgow: Collins ELT, 1990.