

# A coleta de *corpus* de aprendizes: questões qualitativas em uma pesquisa sobre a escrita de aprendizes de língua espanhola

(Collecting learner corpus: qualitative issues in research about the writing ability of Spanish language learners)

Celso Fernando Rocha<sup>1</sup>

<sup>1</sup>Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp/Ibilce)

celsotrad@yahoo.com.br

**Abstract:** In this paper we discuss and present the first results gathered from the compilation of a learner corpus comprised of texts written by university students (Language, Literature and Translation Course). We made use of Corpus Linguistics and observations of researchers from the learner’s corpus field, in order to compile and analyze a corpus of argumentative and descriptive texts written in Spanish. Four hundred compositions were collected (about 120 thousand words) from August 2011 to December 2012. The methodology adopted assisted us considerably in maintaining a comprehensive working agenda, taking into consideration students’ needs and using the data collected as subsidy to improve classroom management of content. We also present the difficulties faced during the data gathering and propose procedures do avoid or minimize them.

**Keywords:** Corpus linguistics; learner corpus; Spanish language; lexicon and pedagogy

**Resumo:** Apresentamos neste artigo os primeiros encaminhamentos adotados durante a coleta e a análise de textos escritos por alunos universitários. Utilizamos a Linguística de Corpus e pontuações de autores que pesquisam a elaboração de corpus de aprendizes, com vistas à construção e à posterior análise de um corpus em espanhol composto de textos argumentativos e descritivos. Foram coletadas 400 redações (aproximadamente 120 mil palavras) no período de agosto de 2011 a dezembro de 2012. A metodologia adotada ajudou consideravelmente no estabelecimento de uma agenda de trabalho docente atenta às necessidades reais dos aprendizes, além de fornecer subsídios para um melhor encaminhamento, em sala de aula, das principais dificuldades apresentadas. Tratamos, também, das dificuldades enfrentadas durante a coleta e propomos procedimentos para evitá-las ou minimizá-las.

**Palavras-chave:** Linguística de Corpus; corpus de aprendizes; língua espanhola; pedagogia e léxico

## Introdução

A construção de *corpora* escritos de aprendizes de língua estrangeira (LE) e seu estudo permitem, entre outros aspectos, sistematizar e entender com que regularidade os alunos empregam recursos gramaticais específicos, as relações entre contextos, a frequência de uso do léxico e os fatores que influenciam diretamente a sua escolha. Desse modo, os *corpora* servem ao propósito de descrição linguística e são úteis para uma prática pedagógica pautada nas necessidades reais e mais recorrentes dos estudantes.

Nesse sentido, anteriormente ao estudo de tais traços, cabe-nos proceder ao delineamento cuidadoso das etapas envolvidas na coleta de um *corpus*, o qual, na maioria das vezes, pode não ser capaz de antecipar com abrangência as instâncias problemáticas

(marginais e determinantes) ao largo da composição dos textos pelos alunos e seu envio ao pesquisador. É necessário, por sua vez, verificar o conhecimento que o aluno (principalmente o aluno de Letras, em contexto brasileiro) traz sobre o aparato tecnológico básico (funcionamento de processadores de texto, correio eletrônico e gerenciamento de arquivos), pois a falta de tal saber pode comprometer sobremaneira as coletas que perdurarão por certo tempo.

Por esse motivo, o presente trabalho tem como foco a apresentação dos passos metodológicos adotados durante a criação de um *corpus* de textos (escritos) de universitários brasileiros aprendizes de língua espanhola. Exploraremos as dificuldades enfrentadas no gerenciamento diário desta compilação e proporemos algumas ações possíveis para contorná-las, uma vez que observamos, nesse contexto, uma lacuna na literatura da área a respeito da descrição do trabalho cotidiano envolvido na produção de *corpora* de aprendizes. Mais especificamente, as obras da área tratam do ensino das ferramentas oferecidas pela linguística de *corpus* com vistas ao desenvolvimento da autonomia do aprendiz, da descrição do material coletado e, de forma abreviada, mencionam que criar *corpora* nesse âmbito exige muito esforço do pesquisador.

Tendo como mote o contexto apresentado, discutiremos alguns temas relacionados à confecção dos textos pelos alunos, ao gerenciamento da coleta do *corpus* e ao tratamento de algumas inadequações mais frequentes nas dissertações levantadas por meio do instrumental oferecido pela Linguística de *Corpus*.

Apresentaremos, em continuação, uma descrição do arcabouço teórico-metodológico empregado neste estudo.

## **Breves definições norteadoras e histórico da área**

O termo *corpus* possui diversas definições para a Linguística de *Corpus*: 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 automática 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.

Nos últimos anos, a presença de *corpora* torna possíveis análises que não teriam grandes possibilidades de ser 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 o posterior levantamento de frequências e contrastes apresentavam restrições de execução, devido ao escopo e à constituição do objeto de estudo.

Para Berber Sardinha (1999, p. 12), 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)

Com relação ao *corpus* de aprendiz, observa-se que, desde o início do século XX, houve experiências que marcaram o ensino de inglês como LE ao serem empregados princípios da linguística de *corpus* na confecção de listagens com o léxico mais frequente desse idioma. Alguns nomes merecem destaque por tal empreitada: Edward Thorndike, Herold Palmer, A. S. Hornby, Irving Lorge e Michael West (cf. BERBER SARDINHA, 2004, p. 252). Os dois primeiros pesquisadores propuseram o *Teacher's Workbook* (1921), que apregoava o ensino do idioma com base na restrição léxica, ou seja, os alunos entrariam em contato com determinado número de palavras primeiro e seriam expostos gradativamente ao léxico menos comum ao longo da aprendizagem. Esse material foi elaborado com base em um levantamento feito a partir de 4,5 milhões de palavras e também foi tomado como modelo para o ensino de língua materna. Palmer e Hornby elaboraram uma *Interim Report on Vocabulary Selection*, que, posteriormente, em 1953, foi assumido por West, que incorporou os trabalhos de Thorndike, Lorge, Palmer e Hornby, criando a *General Service List of English Words*. Tal lista continha as 2 mil palavras mais utilizadas do inglês e suas definições (retiradas de dicionário).

O trabalho desenvolvido pelos investigadores merece destaque, pois se trata de coleta e análise de grande quantidade de dados em uma época em que não existia o computador. Na atualidade, mesmo com os recursos computacionais disponíveis, a análise de milhões de palavras não é tarefa simples.

Na atualidade, destaca-se em língua espanhola o trabalho pioneiro de Soto Balbás (2003), que compila um *corpus* de aproximadamente 25 mil palavras com foco na descrição dos erros mais frequentes em 216 redações de aprendizes de E/LE. Também temos conhecimento de trabalho que investiga a competência tradutória no par linguístico espanhol-português (CINTRÃO, 2009), e Rocha (2013) descreve o uso dos verbos *haber* e *tener* em um *corpus* de aproximadamente 100 mil palavras, organizado a partir de textos produzidos por alunos universitários de um curso de Letras.

É importante lembrar que durante boa parte do século XX o paradigma teórico mentalista (com base nos preceitos da linguística chomskyana) ofuscou grandemente pesquisas com viés observacional. Para os mentalistas, a grande quantidade de dados observacionais gerados por aquelas pesquisas não era desejável, pois seria suficiente a intuição do falante para resolver questionamentos em relação ao emprego da língua. Nos últimos vinte anos, não obstante, os estudos que fazem uso de *corpora* passam a disponibilizar descrições mais precisas sobre inúmeros traços linguísticos, nos seus mais diferentes registros (linguagem oral, discurso formal, discurso acadêmico, diferentes dialetos, etc.). A relevância desse tipo de pesquisa reside no fato de que os resultados podem auxiliar no encaminhamento mais adequado, em sala de aula, de questões relacionadas ao uso de língua, na elaboração de material didático e de currículos mais atentos às necessidades locais. Segundo Berber Sardinha (2004), o desenvolvimento de materiais didáticos a partir da perspectiva de falantes nativos pode produzir visão idealizada da aprendizagem de

uma 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-los 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.

Coletar *corpora* de aprendizes é tarefa árdua e exige muito esforço e tempo inestimável por parte dos pesquisadores (GRANGER, 2002, p. 7); portanto, é imprescindível um plano específico para coleta, armazenagem e extração de dados. Desse modo, a perspectiva da linguística de *corpus* propicia um ponto de observação privilegiado. De acordo com Lavid (2005, p. 336),

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.

A título de ilustração, podemos mencionar estudos em língua inglesa que demonstram a existência de 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 coocorrência com outros itens linguísticos. Expressões que se mostraram mais frequentes em textos de aprendizes (*I believe that*, *In my opinion* e *I am not of the same opinion*) não se apresentaram com frequência significativa em textos de nativos. A hipótese levantada foi a de que haveria certa transferência de padrões típicos da linguagem oral de sala de aula para o texto escrito. Há também estudos que apontam para as diferenças entre o léxico empregado por falantes nativos e o usado por aprendizes, dando especial atenção ao processo de interlíngua.

### **Construção do *Corpus* e passos metodológicos**

O *corpus* está sendo compilado a partir de textos escritos por alunos dos cursos de Letras (Bacharelado e Licenciatura). Trata-se de um *corpus* orgânico que vem sendo alimentado desde junho de 2011 com redações argumentativas e descritivas (temas variados).

Quanto à extensão dos textos, pedimos aos alunos que escrevessem redações com o mínimo de 300 e o máximo de 500 palavras. Quando alguma redação não atingia o número mínimo, entramos em contato com o aluno e discutimos as possibilidades de ampliá-la (procedimento não adotado nas redações do primeiro ano), 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 27 temas propostos, escolhemos seis para serem trabalhados em sala de aula, sem pré-levantamento de informações e sem preparação de rascunho (digitadas no laboratório). Até dezembro de 2012, o *corpus* continha aproximadamente 400 redações (120 mil palavras).

Para as 21 redações restantes, o aluno recebeu o tema do mês com antecedência, e pôde realizar rascunhos em casa. Além disso, foi permitido pesquisar livros, dicionários e sítios na internet, bem como consultar o professor e os colegas. Houve, também, durante as aulas, momentos para a discussão do tema, a contextualização por meio de vídeos e a explicitação dos componentes essenciais requeridos de um texto descritivo ou argumentativo.

Pedimos que o aluno desse um título ao seu texto e o digitasse no processador de texto *Wordpad* (*Microsoft*), salvando o arquivo no formato txt (texto sem formatação ou “documento texto” no Windows 7), com o seguinte formato de nome: “número da redação”, “tema”, “D” para descrição e “A” para argumentação, além das “iniciais de seu nome completo”.

Exemplo: “01PoluiçãoNasGrandesCidades.A.CFR”.

Instruímos o aluno a registrar dentro do arquivo (somente da primeira redação escrita) algumas informações, a saber: “idade”, “sexo”, “há quanto tempo estuda espanhol”, “quanto tempo teve para preparar o rascunho” (em horas), “quais fontes pesquisou” e “quais dificuldades encontrou”.

Depois que o arquivo nos foi enviado, fizeram-se duas cópias dele: uma para ser corrigida e devolvida ao aluno, e outra para fazer parte do *corpus* de estudo. No que concerne à correção, sublinhamos algumas inadequações nos textos (já havíamos inserido correções tradicionais em outros trechos, antes da entrega) e pedimos que encontrassem alternativas corretas. Quando não conseguiam, apresentávamos possíveis soluções.

Neste trabalho, apresentamos uma descrição qualitativa do processo de coleta, relatando as principais dificuldades enfrentadas nos primeiros meses de trabalho em relação ao gerenciamento do *corpus* e discutimos algumas soluções adotadas. Também apresentamos algumas análises focadas no emprego do léxico.

## **Os principais desafios na compilação do *corpus* de aprendizes e os encaminhamentos adotados**

Na literatura da área, encontram-se menções ao fato de que compilar *corpora* de aprendiz é tarefa difícil, que consome um tempo inestimável (Granger, 2002, p. 7), sendo necessária a elaboração de um plano de coleta estruturado e preciso, com vistas à otimização do processo. Nesse ponto, relatos sobre o cotidiano da coleta ou dos problemas mais comuns em contextos específicos não são frequentes, não se descrevem os “bastidores da pesquisa”, o labor diário na compilação de *corpora*. Tal fato é compreensível, pois o foco é a descrição da língua, a elaboração de glossários, entre outros aspectos. No entanto, o conhecimento procedimental é de grande valia e serve para que a comunidade de pesquisadores tenha acesso ao que é feito durante fases de pré-coleta, coleta e pós-coleta (muitas vezes, tais detalhes são mencionados nas conversas informais entre pesquisadores ou se perdem sem documentação).

Nesse sentido, temos como norte apresentar os principais desafios encontrados em nossa coleta e os procedimentos adotados para tornar mais eficiente e eficaz o plano metodológico de coleta previamente elaborado. Primeiro, trataremos de questões relacio-

nadas ao conhecimento que o aluno traz sobre processadores de textos e correio eletrônico; posteriormente, de sua disposição para a escrita e o atendimento de prazos; por fim, forneceremos alguns exemplos relacionados à estruturação do texto em língua espanhola e ao uso de dicionários.

## **O uso do instrumental tecnológico pelos alunos de licenciatura/bacharelado em Letras**

Um dos primeiros pontos que merecem atenção é a verificação do conhecimento de que o aluno dispõe em relação ao ferramental tecnológico envolvido no processo de criação e manipulação de arquivos eletrônicos; mais especificamente, a verificação do que ele sabe sobre processadores de textos, e-mail e transferência de arquivos, uma vez que as redações, na maioria dos casos, são compiladas e circulam nesse meio.

À primeira vista, poderia parecer uma questão irrelevante, dado o contexto de imersão tecnológica em que trabalhamos e vivemos e ao qual boa parte dos alunos da universidade pública do estado de São Paulo tem acesso. O computador ganha cada vez mais espaço na casa do brasileiro e nas universidades, tornando-se ferramenta imprescindível e onipresente. No entanto, há um hiato entre o que se faz com a máquina nas horas de lazer e de estudo e durante a criação estruturada de conhecimento. Essa lacuna levou a uma das primeiras barreiras a serem superadas rumo à compilação de nosso *corpus* de aprendizagens. Vários alunos apresentaram arquivos com organização textual não adequada em processador de texto e os enviaram sem seguir os passos ensinados.

Os principais aspectos que necessitaram de encaminhamento adequado são: a) espaçamento entrelinhas; b) paragrafação; c) margens; d) salvar arquivos, respeitando a nomenclatura oferecida; e) escrita de *e-mails*; f) anexar arquivo. Estes, quando não respeitados, tornam o trabalho de coleta moroso e suscetível ao erro, principalmente quando diversas turmas enviam redações, em pequenos intervalos de tempo.

Antes da impressão das redações com a escrita de comentários manuscritos ou com a correção pelo professor ou pelo aluno, tivemos que reorganizá-las, levando-se em consideração o espaçamento duplo, a paragrafação e as margens acordadas. Esta formação adotada permitiu a visualização e a edição à mão com maior rapidez. Nas primeiras correções, não atentamos com maior acuidade para essa necessidade organizacional, o que gerou lentidão e desorganização no momento da inserção de palavras e da reescrita de parágrafos por conta da refacção. Como havia muitas observações feitas, os alunos tiveram dificuldades para a reescrita dos textos, que seria terminada em casa. Alguns mencionaram que, quando estavam reeditando suas redações, não conseguiam se lembrar das correções feitas e das palavras que teriam que reescrever, seja por falta de notações mais precisas seja por terem escrito sem respeitar as normas estruturais estabelecidas. Assim, começamos a enfatizar a importância dessas normas e, paralelamente, incentivá-los ao uso de estratégias de notação diferenciadas (canetas de cores distintas e notas de rodapé). Essas estratégias foram aplicadas, e, conforme o relato dos alunos, tornou-se mais eficiente o entendimento das notações de revisão feitas em sala.

Como tratado na seção de “metodologia de coleta”, os alunos foram instruídos, desde a primeira redação, a nomear os arquivos e salvá-los em formato “doc” (Microsoft

Office) e, posteriormente, em txt (texto sem formatação). Alguns alunos os entregaram em formato doc, ignorando a etapa de conversão, alegando esquecimento e o não conhecimento de como fazê-lo. Em uma aula de laboratório, instruímo-los a realizar o procedimento e ilustramos com uma apresentação em projetor o “passo-a-passo” da atividade. Nas entregas seguintes, 80% dos textos já se encontravam ajustados em termos de formatação interna e externa. Selecionamos os autores dos 20% que ainda não estavam adequados e oferecemos um tempo para sanar dúvidas.

A forma de nomear o arquivo havia sido estabelecida desde o primeiro dia de aula, no entanto, era a instrução que ainda não encontrava adesão por parte dos discentes. Nas primeiras redações, aproximadamente 90% dos alunos não enviaram os arquivos com os nomes acordados. Muitos escreviam apenas “redação1.doc” ou “redaçãosobreambiente.doc”. No final de 2012, a persistência de desvios (da forma previamente adotada) encontrava-se na faixa dos 30%, e em pelo menos três momentos distintos repassamos o formato que deveriam adotar.

Cabe dizer que o principal problema em relação à não adoção do padrão encontrava-se no fato de que há necessidade de abrir cada arquivo, verificar o tema proposto e, então, decidir a pasta na qual será salvo. A não entrega na data marcada também agravava essa situação, pois as datas de recepção estavam predefinidas para cada grupo. Dessa forma, além de verificar o tema, tínhamos que confirmar o curso (licenciatura ou bacharelado) do qual o aluno fazia parte, evitando, assim, salvar o arquivo na pasta de outro grupo.

A entrega de algumas redações fora de prazo, além de atrapalhar o arquivamento, como mencionado, gerou atrasos nos encaminhamentos em sala de aula.

Os alunos também apresentaram desconhecimento em relação aos dados (de remetente, assunto e identificação de turma) a serem oferecidos em uma mensagem eletrônica. Um dos casos mais comuns foi o envio de mensagens sem nenhuma indicação de assunto ou identificação, como, por exemplo, nome do aluno, turma e tema da redação: algumas eram bem sucintas, “Prof. Meu texto ae!”. Primeiramente, respondíamos o e-mail, solicitando maiores informações. Em ocasião oportuna, explicamos em sala de aula a formatação básica de um e-mail. Retornamos alguns deles sem arquivo anexo, ou com o texto digitado no corpo da mensagem, para que fossem readequados e reenviados.

## **Disposição à escrita e atendimento de prazos**

Um dos fatores que devemos levar em consideração ao compilarmos um *corpus* de aprendiz é a disposição do aluno em cumprir as tarefas pedidas. Nada mais natural quando se trata de escrita. Sabemos que escrever um texto, argumentar e expor ideias de forma coerente e coesa é uma tarefa complexa e laboriosa, sujeita à ativação de conhecimentos prévios e dependente de repertório linguístico individualizado. Quando se trata de língua estrangeira, mais especificamente do espanhol, que guarda algumas similaridades (e inúmeras diferenças) com o português, o trabalho se torna intenso e dependente de disponibilidade de tempo extra para verificar as opções de expressão escrita mais adequadas.

Assim, o tópico “disponibilidade de tempo” foi o mais mencionado pelos alunos ao justificarem a não entrega. Pudemos constatar, após várias conversas, que trazem consigo uma visão de tempo distorcida e mais atrelada a conceitos puramente psicológicos,

isto é, dispõem de tempo, mas acabam, por falta de organização pessoal, postergando inúmeras atividades e entrando em um ciclo de ansiedade e frustração. Pequenos “surto” eclodem no momento em que há um prazo expirado ou cobrança pelo docente. Estas situações encontraram “terreno fértil” no momento em que fomos mais compreensíveis e flexíveis em relação às datas, ou seja, houve mais questionamentos e indagações sobre entrega, conteúdo, quantidade de palavras e necessidade de escrita. Alguns alunos começaram a argumentar que já escreviam na disciplina Prática de Tradução em Língua Espanhola e não precisariam escrever mais. Paralelamente a essa questão, chamou-nos atenção a ideia equivocada de que os textos são construídos de forma não laboriosa e sem necessidade de revisões.

Em determinada ocasião, quando devolvíamos os textos sublinhados, pudemos observar vários alunos dobrando a folha de papel e guardando dentro dos cadernos, com pouca atenção dada ao fato de que havia somente palavras sublinhadas e poucas inserções de correção. A partir desse momento, abrimos um espaço para discutir brevemente a concepção de construção de texto que acolham (para a Linguística, a noção de texto ainda está aberta). Pensavam em texto como um produto dependente do tempo e da disposição a escrever, uma concepção mercantilizada e individualista. A nota ainda aparecia como a motivação principal para a escrita.

Uma vez que, no início da coleta, por volta de 30% dos alunos não entregaram os textos pedidos ou entregaram com certo atraso, optamos por abrir um espaço maior em sala de aula para a conscientização sobre os benefícios da feitura dos textos e das discussões prévias dos temas propostos. Para tanto, empregamos vídeos, textos, discussões individualizadas e coletivas de forma mais intensa. Também explicamos os benefícios da escrita quinzenal, o *feedback* pontual do professor, acompanhando o desenvolvimento individual do aluno e oferecendo aulas com base nas dificuldades mais frequentes.

Na discussão em sala de aula, outras questões foram incluídas: a consulta a fontes externas; o repertório; a criação de rascunhos; o tempo de consulta e de elaboração da redação, e a releitura. Começamos a enfatizar também o conceito de escrita da redação como um processo de estabelecimento de vínculo duradouro. O texto (durante e após a escrita) torna-se parte integrante do desenvolvimento pedagógico do aluno. Continua como instrumento de mudança e de aprendizagem ao oferecer uma janela para verificação do *status quo*, ou seja, em que estágio de desenvolvimento da escrita o discente se encontra. Essa mudança de paradigma foi conseguida por meio de discussões e da instituição de um portfólio. Foram oferecidas algumas pastas (etiquetadas e identificadas com o nome do aluno e a turma) para que as redações fossem arquivadas e consultadas com maior rapidez. Além da consulta, o aluno iria escolher uma delas, no final do ano, para ser lida na sala e receber uma nota. Todo o portfólio, por sua vez, seria avaliado em termos de organização e de assiduidade na guarda dos textos. Nota-se que há um equilíbrio entre a conscientização da necessidade de valorização da escrita em prol do desenvolvimento do educando e a elaboração de instrumentos de controle avaliativos.

No final de 2012, por volta de 10% dos alunos continuavam não entregando as redações. As explicações mais comuns foram: a) falta de tempo; b) não estar presente no dia da discussão do tópico e da apresentação de material audiovisual; c) esquecimento e pedido de novo prazo e d) silêncio. Algumas redações também não atingiram o número mínimo de palavras estabelecido, sendo devolvidas e reescritas pelo aluno, a partir de algumas notações indicando os tópicos que poderiam ser mais bem desenvolvidos.



Explicitações constantes sobre o número mínimo de palavras, avisos sobre as datas previstas para a entrega e a introdução do portfólio foram suficientes para mitigar os problemas relatados.

## **A estruturação do texto, o uso de dicionários e encaminhamentos**

Nosso intuito, ao propor a coleta do *corpus* de aprendiz, além de descrever o léxico mais frequente, foi oferecer subsídios para que os alunos melhorassem a produção escrita em língua espanhola. Assim, mantivemos um diário, com notações qualitativas, indicando o nome dos alunos e os parágrafos que necessitavam de revisão. Em sala, após a correção de cada tema proposto, apresentávamos um resumo qualitativo de todas as redações. Cada aluno recebia a sua redação com indicações (sublinhados para pesquisa posterior), inserções ou comentários.

Dessa forma, logo nos primeiros textos entregues, começamos a perceber a necessidade de introdução de conteúdo relacionado à estruturação das redações: noções sobre parágrafo, parágrafo de transição, encadeamento lógico e progressão de ideias. O excerto 1, a seguir, retirado de uma redação<sup>1</sup> sobre o meio ambiente, ilustra nossa colocação:

*El ser humano es un mamífero, que como todos los otros, posee un alto nivel de complejidad, pues cada órgano tiene interrelaciones con los otros sistemas. Sin duda, el ser humano es el animal más distinto de toda la naturaleza, pero aun así, el hombre continua siendo una obra de la naturaleza, aunque algunos no aceptan esto.*

*El medio ambiente, criado para nuestro uso, pero la globalización, la ambición del hombre ha colocado todo a perder. El buraco en la capa de ozono, las quemadas, el efecto estufa entre otros contribuyen con las causas de la autodestrucción humana. (Excerto 1 - redação 13 - sobre o meio ambiente – aluno do segundo ano)*

No primeiro parágrafo (introdutório), o aluno começa a discorrer sobre a complexidade do ser humano e, de modo vago, menciona que cada órgão estabelece relações com outros sistemas (quais?). Na sequência, observa-se uma contradição, pois fora mencionado, primeiramente, que o homem é mais um animal entre os mamíferos e, posteriormente, nega-se tal fato, afirmando sua singularidade (é o animal mais diferente da natureza). Não há conexão entre o primeiro parágrafo e o segundo. Neste, observamos vários tópicos relacionados ao meio ambiente, porém, dispersos e não retomados no restante da redação. Há mais uma incoerência no final do segundo parágrafo: a ideia de que a “autodestruição humana” possa ser causada por uma fonte externa. Talvez a “autodestruição da espécie humana”, por fatores criados por ela mesma, seria uma colocação menos problemática. Podemos destacar algumas questões lexicais que foram pesquisadas e encaminhadas, por meio de anotações, ao aluno. Uma delas se refere ao uso da expressão *continúa siendo*, que, além de transmitir uma ideia de que o homem não deveria fazer parte da natureza e, em algum momento, deixará de fazer, não é tão comum quanto *sigue siendo*. A primeira forma registra aproximadamente 600 casos no *Corpus de Referencia del Español Actual* (CREA), e a segunda, 4.186 ocorrências. Na web, o uso mais frequente é de *sigue siendo*

<sup>1</sup> Apresentamos o primeiro e o segundo parágrafos de uma redação sobre a destruição do meio ambiente provocada pela ação humana. Os alunos assistiram a vídeos, leram textos e discutiram em sala de aula algumas questões pontuais sobre o assunto. O texto encontra-se sem correções.

(24 milhões de ocorrências) em comparação com *continúa siendo* (3 milhões). Em língua portuguesa, há inversão da preferência de uso, “continua sendo” registra 3,2 milhões de usos, e “segue sendo”, 400 mil.

Entre outras questões destacadas no texto, também podemos comentar o uso de *colocando todo a perder*. Nesse caso específico, instruímos o aluno a exemplificar o que seria perdido e a levar em consideração o uso desta expressão em seu texto. Primeiro, pedimos que fosse consultado o verbo “colocar” em dicionários espanhóis monolíngues, devido ao fato de o aluno ter mencionado o uso de um dicionário bilíngue. Os resultados apresentados foram, resumidamente:

- (01) Colocar algo em um lugar que lhe corresponda. (*Colocar el libro en el estante, los puntos sobre las íes*).
- (02) Lançar coisas (*Colocó la bala en el mismo centro*).
- (03) Fazer com que determinada coisa, pessoa, etc. assumira certa posição. (*Coloca la cabeza alta*).
- (04) Proporcionar um emprego ou uma colocação profissional a alguém. (*ha colocado las tres hijas*).
- (05) Com *bien* ou advérbio correspondente (transitivo) casar alguém. (*ha colocado bien a las tres hijas*).
- (06) Fazer com que alguém escute uma história chata. (*te colocará el cuento de siempre*).
- (07) Produzir (uma droga) efeitos alucinatórios em alguém. (*opio y marihuana son drogas que colocan*).

Sugerimos, em seguida, que se pesquisasse (no CREA e na web) o uso das expressões: *colocando todo a perder*, *colocar todo a perder*, *poniendo todo a perder*, *poner todo a perder*, *echando todo a perder* e  *echar todo a perder*.

Os resultados foram:

**Tabela 1. Busca do uso das expressões com o colocado *todo a perder***

<i>cluster</i>	CREA – número de ocorrências	Web – número de ocorrências
<i>colocando todo a perder</i>	Nenhuma	Nenhuma
<i>colocar todo a perder</i>	Nenhuma	15
<i>poniendo todo a perder</i>	Nenhuma	4
<i>poner todo a perder</i>	Nenhuma	31
<i>echando todo a perder</i>	Nenhuma	120 mil
<i> echar todo a perder</i>	3	263 mil

Verifica-se, por meio da Tabela 1, que há preferência pelo verbo  *echar* na construção analisada, e que o uso feito pelo aluno do verbo “colocar” não encontra nenhuma ocorrência nos *corpora* de apoio pesquisados.

Além dessas duas expressões apresentadas, o aluno foi convidado a buscar as diferenças (ou a existência) de outras expressões de seu texto, como, por exemplo: a) *efecto*

*invernadero x efecto estufa, b) buraco en la capa de ozono x agujero en capa de ozono, c) tienen interrelaciones x establecen inter-relaciones.*

Por meio dos dados levantados, houve conscientização dos alunos sobre a importância do uso de dicionários bilíngues e monolíngues. A turma comentou que, muitas vezes, utilizava o bilíngue por ser “mais prático”, esquecendo-se de verificar nos monolíngues e na web (ou no CREA) as possibilidades de usos mais frequentes. Outro aspecto comum é – ao efetuarem buscas nestas obras – selecionarem a primeira definição do verbete, sem preocupação com as diversas acepções apresentadas. A questão do tempo “longo” de leitura de entradas, aliada à concepção equivocada de que deter-se nelas não seja tão importante, pode contribuir para o emprego impreciso do léxico. Nesse sentido, começamos a romper com essa prática com exemplos de usos não frequentes, baseados no *corpus* coletado. Os exemplos evidenciam precipitações na hora de consulta, pois, muitas vezes, a primeira definição é empregada, sem maiores preocupações com o contexto.

Cabe salientar que, de modo análogo, o mesmo procedimento de busca solicitado ao aluno, autor do excerto, foi empregado sistematicamente em outras redações das turmas.

## **Observações finais**

As dificuldades enfrentadas ao longo do processo de coleta do *corpus* propiciaram momentos de reorganização do plano inicial de compilação e de elaboração de estratégias de ensino mais abrangentes e atentas às necessidades dos estudantes dos cursos de Letras (licenciatura e bacharelado). Incluímos, com mais frequência, aulas sobre a organização textual, o uso de ferramentas computacionais, as formas de notação e de refacção textual, fazendo uso da língua meta, ou seja, passamos a oferecer conteúdo específico (técnico), por meio da LE.

Nesse sentido, atrelar a pesquisa à verificação das injunções educacionais locais, sistematizando as dificuldades pontuais (em termos de conhecimento do aluno sobre texto escrito, tecnologia e comportamento diante de prazos) e buscando a aprendizagem efetiva do discente, foi salutar. O espaço criado serviu para ampliar consideravelmente a abrangência do processo pedagógico. Os alunos começaram a assumir comportamentos mais eficientes e eficazes em termos de organização pessoal, os prazos passaram a ser respeitados, uma vez que houve compreensão profunda a respeito dos benefícios do texto escrito no processo de aprendizagem/aquisição de LE. Anteriormente, eles equacionavam (precarosamente) a feitura do texto ao recebimento de nota, sem qualquer explicitação de questões pedagógicas subjacentes ou dos benefícios advindos de tal procedimento.

No que tange à manutenção dos comportamentos e das novas atitudes citadas, cabe dizer que a repetição das instruções, por parte do docente, conferiu maior estabilidade aos processos envolvidos na coleta e na refacção dos textos.

Por fim, esperamos que as colocações apresentadas neste artigo possam ser úteis aos pesquisadores que venham a trabalhar com *corpus* de aprendizes.

## 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.
- GRANGER, S. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam/Philadelphia: John Benjamins, 2002.
- LAVID, J. *Lenguaje y nuevas tecnologías: nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid: Cátedra, 2005.
- 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.
- ROCHA, C. F. O léxico em textos de aprendizes de língua espanhola: um estudo descritivo pautado na linguística de corpus como subsídio para o ensino. *Estudos Linguísticos*, São Paulo, v. 1, n. 40, p. 284-300, 2013.
- 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.
- SOTO BALBÁS, M. *Análise de erros, baseada na Linguística 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.