

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS**

ANA MARÍA FRITZ HERRERA

**ESTUDO CONTRASTIVO DA INTERLÍNGUA EM *CORPUS ORAL E ESCRITO DE*
*APRENDIZES DE ELE***

UBERLÂNDIA

2016

ANA MARÍA FRITZ HERRERA

ESTUDO CONTRASTIVO DA INTERLÍNGUA EM *CORPUS ORAL E ESCRITO DE APRENDIZES DE ELE*

Texto de dissertação apresentado ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Letras e Linguísticas da Universidade Federal de Uberlândia (ILEEL/UFU).

Orientador: Prof. Dr. Ariel Novodvorski

FINALIDADE: Desenvolvimento de uma investigação em nível de Mestrado, com enfoque analítico-descritivo, em relação a aspectos de interlíngua, num *corpus* oral e escrito de aprendizes de espanhol como língua estrangeira.

UBERLÂNDIA

2016

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

H565e Herrera, Ana María Fritz, 1987-
2016 Estudo contrastivo da interlíngua em corpus oral e escrito de
aprendizes de ele / Ana María Fritz Herrera. - 2016.
0 f.

Orientador: Ariel Novodvorski.
Dissertação (mestrado) - Universidade Federal de Uberlândia,
Programa de Pós-Graduação em Linguística.
Inclui bibliografia.

1. Linguística - Teses. 2. Interlíngua (Aprendizagem de línguas) -
Teses. 3. Aquisição da segunda linguagem - Teses. 4. Língua espanhola -
Vícios de linguagem - Teses. I. Novodvorski, Ariel. II. Universidade
Federal de Uberlândia. Programa de Pós-Graduação em Linguística. III.
Título.

AGRADECIMENTOS

Primeiramente quero agradecer ao meu Orientador, o Professor Doutor Ariel Novodvorski, por ser meu principal suporte acadêmico, quem me guiou pelos caminhos da Linguística de *Corpus* e dos Estudos Contrastivos de Interlíngua, traspassando também por aprendizados de vida muito valiosos, por exemplo, saber afrontar situações complexas de vida em um país estrangeiro e saber ser professional ante tudo. Agradeço pela dedicação, pelo carinho e pela compreensão.

Aos amigos que conheci em Uberlândia, os quais foram uma benção, minha família que escolhi no estrangeiro e que conservarão sempre um espaço especial na minha vida. Obrigada pelo infinito apoio e pelo afeto, pois foi fundamental para concluir essa etapa.

À minha mãe que, a pesar da distância soube estar, presente. Sua imagem e seu apoio sempre me inspirou para me superar na vida e sempre será minha fortaleza em todo momento.

Agradecer também a OEA, no contexto do Programa GCUB Universidades Brasileiras, que fizeram possível poder cursar o Programa de Pós Graduação de Mestrado em Estudos Linguísticos.

RESUMO

O presente trabalho tem como objetivo fazer uma representação dos “erros” produzidos pelos aprendizes de espanhol do Curso de Letras/Habilitação em Espanhol da Universidade Federal de Uberlândia. Para esse fim, foi compilado um *corpus* linguístico a partir de produções orais e escritas dos alunos do segundo, quarto, sexto e oitavo período. Os principais temas e autores que deram sustento teórico ao nosso estudo, quanto às análises descritivas, foram: Interlíngua (CORDER, 1967; SELINKER, 1972; BARALO, 1999, 2004; DURÃO, 2007), Linguística Contrastiva (SÖHRMAN, 2007), Modelo para Análise de Erros (DURÃO, 2004; ANDRADE, 2011; SANTOS GARGALLO, 2004), dentre os principais. Cabe destacar que adotamos uma perspectiva de análise de base empírica, apoiados nos subsídios que propicia a Linguística de *Corpus* (BERBER SARDINHA, 2004). Outro componente importante nessa dissertação foi a metodologia. Ela está detalhada passo a passo desde o levantamento e leitura do referencial teórico, até a finalização da escrita do trabalho. Presentando-se, dessa maneira, como um futuro referencial para pesquisas que se baseiam na utilização de LC como abordagem metodológica e na análise de erros de aprendizes. As análises desenvolvidas no decorrer desse trabalho, compreenderam primeiramente o dimensionamento dos *corpora* utilizados, seguido de listas das palavras mais recorrentes, análises quantitativas e qualitativas, as quais constituíram um mapeamento dos “erros”, outorgando, desta maneira, um valor potencial ao se tratar de um estudo que poderá ser utilizado como referente para uma eventual elaboração de material didático, pensado especialmente para as aulas de espanhol ofertadas no Curso de Letras/Habilitação em Espanhol da Universidade Federal de Uberlândia.

Palavras-chave: Interlíngua. Estudo Contrastivo. Erros. Curso de Letras/Habilitação Espanhol.

RESUMEN

El presente trabajo tiene como objetivo hacer una representación de los “errores” producidos por los aprendices de español del Curso de Letras (Habilitación en Español de la Universidad Federal de Uberlândia. Para este fin, fue compilado un *corpus* lingüístico a partir de las producciones orales y escritos de los alumnos del segundo, cuarto, sexto y octavo semestre. Los principales temas y autores que dieron sustento teórico a nuestro estudio, en cuanto a los análisis descriptivos fueron: Interlengua (CORDER, 1967; SELINKER, 1972; BARALO, 1999, 2004; DURAO, 2007), Lingüística Contrastiva (SÖHRMAN, 2007), Modelo para Análisis de Errores (DURAO, 2004; ANDRADE, 2011; SANTOS GARGALLO, 2004), entre los principales. Cabe destacar que adoptamos una perspectiva de análisis de base empírica, apoyados en los subsidios que propicia la Lingüística de *Corpus* (BERBER SARDINHA, 2004). Otro componente importante en esta tesis fue la metodología. Se detalla paso a paso desde el levantamiento y lectura del referencial teórico, hasta la finalización del proceso de escritura del trabajo. Presentándose de esta manera como un futuro referencial para investigaciones que se basan en la utilización de LC como abordaje metodológico, y en el análisis de errores de aprendices. Los análisis desarrollados en el transcurso de este trabajo, comprendieron primeramente el dimensionamiento de los *corpora* utilizados, seguido de listas de las palabras más recurrentes, análisis cuantitativos y cualitativos, los cuales constituyeron un mapeo de los “errores”, otorgando de esta manera, un valor potencial al tratarse de un estudio que podrá ser utilizado como referente para una eventual elaboración de material didáctico, pensado especialmente para las clases de español que ofrece el Curso de Letras/Habilitación en Español de la Universidad Federal de Uberlândia.

Palabras clave: Interlíngua. Estudio Contrastivo. Errores. Curso de Letras/Habilitación Español.

LISTA DE ILUSTRAÇÕES

Figura 1	Interlíngua	29
Figura 2	Horas de estudo de língua espanhola por ano cursado no Curso de Letras da UFU.	45
Figura 3	Armazenamento na plataforma Dropbox.	46
Figura 4	Scanner Planetário. Biblioteca da UFU, Campus Santa Mônica.	47
Figura 5	Amostra de digitalização das atividades de produção escrita.	48
Figura 6	Armazenamento do <i>corpus</i> escrito na plataforma Dropbox.....	49
Figura 7	Amostra da transcrição da atividade oral de Debate.	50
Figura 8	Amostra do enriquecimento da transcrição ortográfica na atividade de Debate no sexto período.	51
Figura 9	Armazenamento da atividade oral Apresentação, segundo período, no Dropbox.	52
Figura 10	Dimensão do <i>corpus</i> no <i>WordList</i>	53
Figura 11	Dados estatísticos do <i>corpus</i> oral e escrito no WST.	54
Figura 12	Amostra da identificação da categoria grammatical dos erros no quarto período do <i>corpus</i> oral.....	57
Figura 13	Amostra da dimensão do <i>subcorpora</i> do quarto período do <i>corpus</i> oral.....	58
Figura 14	Palavras mais recorrentes na atividade de opinião do <i>corpus</i> escrito.	59
Figura 15	Amostra da tabela com as palavras lexicais mais recorrentes.....	60
Figura 16	Amostra do gráfico de análise quantitativa.	60
Figura 17	Amostra das linhas de concordância.	61
Figura 18	Amostra do formato utilizado para a análise qualitativa.....	62
Figura 19	Listas de palavras mais frequentes no <i>corpus</i> oral e escrito.	64
Figura 20	Lista de palavras por ordem alfabética no <i>corpus</i> oral e escrito.	65
Figura 21	Linhos de concordância pela categoria FON do <i>corpus</i> oral.	66
Figura 22	Linhos de concordância pela categoria MOR do <i>corpus</i> oral.	67
Figura 23	Linhos de concordância pela categoria SIN do <i>corpus</i> oral.	68
Figura 24	Linhos de concordância pela categoria LEX do <i>corpus</i> oral.....	68
Figura 25	Linhos de concordância pela categoria ORT do <i>corpus</i> escrito.	69
Figura 26	Linhos de concordância pela categoria MOR do <i>corpus</i> escrito.	70
Figura 27	Linhos de concordância pela categoria SIN do <i>corpus</i> escrito.	70
Figura 28	Linhos de concordância pela categoria LEX do <i>corpus</i> escrito.	71
Figura 29	Extensão do <i>corpus</i> oral.	73

Figura 30	Extensão do <i>corpus</i> escrito.....	74
Figura 31	Amostra da extensão do <i>corpus</i> oral no segundo período.....	75
Figura 32	Amostra dos dados estatísticos do <i>corpus</i> oral no quarto período.....	75
Figura 33	Amostra dos dados estatísticos do <i>corpus</i> oral no sexto período.....	76
Figura 34	Amostra dos dados estatísticos do <i>corpus</i> oral no oitavo período.....	76
Figura 35	Dados estatísticos da atividade “apresentação”, recorte do <i>corpus</i> oral.....	78
Figura 36	Dados estatísticos da atividade “debate”, recorte do <i>corpus</i> oral.....	78
Figura 37	Dados estatísticos da atividade “diálogo”, recorte do <i>corpus</i> oral.....	79
Figura 38	Dados estatísticos da atividade “leitura”, recorte do <i>corpus</i> oral.....	79
Figura 39	Palavras mais recorrentes do <i>corpus</i> oral e do <i>corpus</i> escrito.....	81
Figura 40	Palavras mais recorrentes no segundo e quarto período no <i>corpus</i> oral.....	83
Figura 41	Palavras mais recorrentes no sexto e oitavo período no <i>corpus</i> oral.....	85
Figura 42	Palavras mais recorrentes nas atividades de apresentação e debate.....	87
Figura 43	Palavras mais recorrentes nas atividades de diálogo e leitura.....	89
Figura 44	Palavras mais recorrentes no segundo e quarto período no <i>corpus</i> escrito.....	91
Figura 45	Palavras mais recorrentes no sexto e oitavo período no <i>corpus</i> escrito.....	93
Figura 46	Palavras mais recorrentes nas atividades de carta e descrição.....	96
Figura 47	Palavras mais recorrentes na atividade de opinião.....	98
Figura 48	Linhos de concordância do segundo período para o tipo de “erro” fonológico...	100
Figura 49	Linhos de concordância da atividade “debate” para o tipo de erro léxico.....	100
Figura 50	Quantitativo dos erros morfológicos por período do <i>corpus</i> oral.....	102
Figura 51	Quantitativo dos erros fonológicos por período do <i>corpus</i> oral. Fonte: a autora.	103
Figura 52	Quantitativo dos erros lexicais por período do <i>corpus</i> oral.....	104
Figura 53	Quantitativo dos erros sintáticos por período do <i>corpus</i> oral.....	105
Figura 54	Quantitativo dos erros morfológicos por atividade do <i>corpus</i> oral.....	107
Figura 55	Quantitativo dos erros fonológicos por atividade do <i>corpus</i> oral.....	108
Figura 56	Quantitativo dos erros lexicais por atividade do <i>corpus</i> oral.....	109
Figura 57	Quantitativo dos erros sintáticos por atividade do <i>corpus</i> oral.....	110
Figura 58	Quantitativo dos erros morfológicos por período do <i>corpus</i> escrito.....	112
Figura 59	Quantitativo dos erros ortográficos por período do <i>corpus</i> escrito.....	113
Figura 60	Quantitativo dos erros lexicais por período do <i>corpus</i> escrito.....	114
Figura 61	Quantitativo dos erros sintáticos por período do <i>corpus</i> escrito.....	115
Figura 62	Quantitativo dos erros morfológicos por atividade do <i>corpus</i> escrito.....	117
Figura 63	Quantitativo dos erros ortográficos por atividade do <i>corpus</i> escrito.....	118

Figura 64 Quantitativo dos erros lexicais por atividade do <i>corpus</i> escrito.....	119
Figura 65 Quantitativo dos erros sintáticos por atividade do <i>corpus</i> escrito.....	120

LISTA DE TABELAS

Tabela 1	Conceitos de lapsos, falhas e erros	23
Tabela 2	Horas de estudo de disciplina especializadas em espanhol durante o curso.....	44
Tabela 3	Extensão do <i>corpus</i> de estudo.....	53
Tabela 4	Códigos de transcrição.....	55
Tabela 5	Níveis de análise das inadequações.....	56
Tabela 6	Códigos para a identificação das categorias gramaticais dos “erros” do <i>corpus</i> linguístico.	58
Tabela 7	Extensão do recorte do <i>corpus</i> de estudo para as análises preliminares.....	65
Tabela 8	Ocorrências de erros em <i>corpus</i> oral e escrito.....	72
Tabela 9	Extensão do <i>corpus</i> de estudo.....	74
Tabela 10	Resumo dos dados estatísticos do <i>corpus</i> oral por períodos.....	77
Tabela 11	Resumo dos dados estatísticos do <i>corpus</i> oral por atividades	80
Tabela 12	Palavras lexicais mais recorrentes no <i>corpus</i> oral	82
Tabela 13	Palavras lexicais mais recorrentes no <i>corpus</i> escrito.....	82
Tabela 14	Palavras lexicais mais recorrentes do segundo período no <i>corpus</i> oral.....	83
Tabela 15	Palavras lexicais mais recorrentes do quarto período no <i>corpus</i> oral.....	84
Tabela 16	Palavras lexicais mais recorrentes do sexto período no <i>corpus</i> oral	85
Tabela 17	Palavras lexicais mais recorrentes do oitavo período no <i>corpus</i> oral	86
Tabela 18	Palavras lexicais mais recorrentes por período no <i>corpus</i> oral.....	86
Tabela 19	Palavras lexicais mais recorrentes nas atividade de “apresentação”	88
Tabela 20	Palavras lexicais mais recorrentes nas atividade de “debate”	88
Tabela 21	Palavras lexicais mais recorrentes nas atividade de “diálogo”	89
Tabela 22	Palavras lexicais mais recorrentes nas atividade de “leitura”	90
Tabela 23	Palavras mais recorrentes por atividade no <i>corpus</i> oral	90
Tabela 24	Palavras lexicais mais recorrentes no segundo período do <i>corpus</i> escrito.....	92
Tabela 25	Palavras lexicais mais recorrentes no quarto período do <i>corpus</i> escrito	92
Tabela 26	Palavras lexicais mais recorrentes no sexto período do <i>corpus</i> escrito	94
Tabela 27	Palavras lexicais mais recorrentes no oitavo período do <i>corpus</i> escrito	94
Tabela 28	Palavras mais recorrentes por período no <i>corpus</i> escrito	95
Tabela 29	Palavras lexicais mais recorrentes na atividade “carta” do <i>corpus</i> escrito.....	96
Tabela 30	Palavras lexicais mais recorrentes na atividade “descrição” do <i>corpus</i> escrito.....	97
Tabela 31	Palavras lexicais mais recorrentes na atividade “opinião” do <i>corpus</i> escrito.....	98

Tabela 32	Palavras mais recorrentes por atividade no <i>corpus</i> escrito.	99
Tabela 33	Quantitativo das ocorrências por período e por tipo de erro do <i>corpus</i> oral.....	101
Tabela 34	Quantitativo das ocorrências por atividade e por tipo de erro do <i>corpus</i> oral. ...	106
Tabela 35	Quantitativo das ocorrências por período e por tipo de erro do <i>corpus</i> escrito. ..	111
Tabela 36	Quantitativo das ocorrências por atividade e por tipo de erro do <i>corpus</i>	116

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Linguística Contrastiva	19
2.2	Conceitos de “Erro” e o Modelo de Análise de Erros	21
2.3	Interlíngua	28
2.4	Linguística de Corpus	30
3	METODOLOGIA E CORPUS	35
4	ANÁLISE DE DADOS	63
4.1	Introdução	63
4.2	Análises Preliminares	63
4.3	Extensão dos subcorpora por período	72
4.4	Extensão dos subcorpora por atividade	78
4.5	Palavras mais recorrentes no corpus oral e escrito	80
4.5.1	<i>Palavras mais recorrentes no corpus oral por período</i>	83
4.5.2	<i>Palavras mais recorrentes no corpus oral por atividade</i>	87
4.5.3	<i>Palavras mais recorrentes no corpus escrito por período</i>	91
4.5.4	<i>Palavras mais recorrentes no corpus escrito por atividade</i>	95
4.5.5	<i>Quantitativo dos “erros” por períodos no corpus oral</i>	99
4.5.6	<i>Quantitativo dos “erros” por períodos no corpus escrito</i>	110
4.6	Estudo qualitativo dos erros	120
4.6.1	<i>Estudo qualitativo dos erros no corpus oral por períodos</i>	120
4.6.2	<i>Estudo qualitativo dos erros no corpus oral por períodos</i>	135
5	CONSIDERAÇÕES FINAIS	147
6	REFERÊNCIAS BIBLIOGRÁFICAS	150
	APÊNDICE A	153
	APÊNDICE B	154
	ANEXO A	155

1 INTRODUÇÃO

Esta pesquisa tem uma motivação que surgiu, num primeiro momento, de uma curiosidade pessoal, como falante nativa de espanhol, com relação à língua espanhola no Brasil. Diversas questões e dúvidas começaram a surgir, a partir da minha chegada a este país, por ocasião do Mestrado em Estudos Linguísticos. Algumas dessas questões, mais relacionadas à aquisição da língua espanhola pelos brasileiros, apontavam ao momento e espaço em que ocorreria esse primeiro contato com a língua: no ensino fundamental, médio ou apenas em cursos livres de idiomas? Por outro lado, a curiosidade por saber se existiria um real interesse pela língua espanhola neste país também surgiu como um assunto instigante, assim como se o espanhol seria uma língua respeitada e falada pelos brasileiros.

Um problema na aquisição do espanhol como língua estrangeira pode ser observado na falta de material didático especializado para falantes nativos de português brasileiro, posto que, na maioria das vezes, o que acontece com esse tipo de material é que tenta abranger o máximo possível quanto à variedade de línguas maternas que possam existir simultaneamente em uma mesma sala de aula, deixando para trás, então, a efetividade que o material poderia chegar a atingir, se fosse especializado para falantes de uma língua só. Como em Durão (2004, p. 18):

Não é a mesma coisa ensinar espanhol para falantes de uma língua tão próxima como é a portuguesa do que fazê-lo com outros falantes de línguas estruturalmente mais afastadas, alemães, árabes ou japoneses, por exemplo.¹

No caso específico de falantes nativos de português aprendizes de espanhol, outro problema identificável é a proximidade das duas línguas, os mecanismos que coincidem são muitos, sendo esta a maior dificuldade. Segundo Durão (2014, p. 18):

Ao mesmo tempo, é frequente na aprendizagem do espanhol por parte dos luso-falantes erros inexplicáveis pela simplicidade das estruturas afetadas, mas que simplesmente são um resultado perante fenômenos que parecem muito semelhantes com a língua própria e que fazem surgir a insegurança acerca da validade do seu espanhol.²

1“No es lo mismo enseñar español a hablantes de una lengua tan próxima como es la portuguesa que hacerlo a otros hablantes de lenguas estructuralmente más alejadas, alemanes, árabes o japoneses, por ejemplo”.

2“Al mismo tiempo son frecuentes en el aprendizaje del español por parte de luso-hablantes equivocaciones inexplicables por la sencillez de las estructuras afectadas, pero que simplemente resultan del recelo ante fenómenos que parecen demasiado semejantes a las de la lengua propia y que hacen surgir la desconfianza sobre su validez en español”.

Um outro motivo instigante para desenvolver o presente estudo foi a lei 11.161 aprovada no ano 2005. A lei anteriormente mencionada trata sobre a inclusão obrigatória da disciplina de língua espanhola no currículo das escolas brasileiras, e que o prazo para concluir essa implementação no currículo seria de cinco anos. Até o ano 2016, pode se apreciar que ainda não tem sido implementada do modo como foi formulada e aprovada faz onze anos.

Várias começaram a ser as dúvidas que me motivavam a desenvolver algum tipo de investigação a respeito da língua espanhola neste país. Contudo, tentando dar uma direção mais clara e específica a essa pesquisa, decidi focar o presente estudo partindo da delimitação do objeto a aspectos característicos da interlíngua, num *corpus* de aprendizes de espanhol como língua estrangeira, sob uma abordagem analítico-descritiva, com subsídios da Linguística de *Corpus* (LC). Nesse sentido, a motivação principal conduziria à identificação e análise dos principais “erros”, produzidos num *corpus* oral e escrito e ao desenvolvimento de uma metodologia restrita, para o estudo dos fenômenos interlingüísticos e de suas prováveis causas.

Conforme a explicação supracitada e pensando também na viabilidade de desenvolvimento de um projeto dessa natureza, delimitamos o grupo de informantes aos estudantes, professores em formação, do Curso de Letras/Espanhol, do Instituto de Letras e Linguística da Universidade Federal de Uberlândia (ILEEL/UFU), levando em consideração os períodos do segundo semestre letivo (segundo, quarto, sexto e oitavo períodos) de 2014.

Para o seguinte estudo, a fundamentação teórica que dará principal sustento às análises descritivas é: Interlíngua (CORDER, 1967; SELINKER, 1972; BARALO, 1999, 2004; DURÃO, 2007), Linguística Contrastiva (SÖHRMAN, 2007), Modelo para Análise de Erros (DURÃO, 2004; ANDRADE, 2011; SANTOS GARGALLO, 2004), dentre os principais. Adotamos uma perspectiva de análise de base empírica, apoiados nos subsídios que propicia a Linguística de *Corpus* (BERBER SARDINHA, 2004), entendida aqui como metodologia e abordagem para a pesquisa. Segundo Parodi (2010, p. 14):

Ora bem, para começar afirmo que a LC nesta versão constitui uma abordagem metodológica para o estudo de línguas e apresenta oportunidades revolucionárias para a descrição, análise e ensino de discursos de todos os tipos. Ele também fornece uma base empírica para o desenvolvimento de materiais educativos e metodológicos de vários tipos, bem como para a construção de gramáticas, dicionários e outros, tanto geral e discursos especializados, orais e escritos. Deste ponto de vista argumento que a LC constitui um conjunto ou coleção de princípios metodológicos para estudar qualquer domínio linguístico e é caracterizada por fornecer apoio à investigação da linguagem em uso a partir de *corpus* linguístico com substrato

em tecnologia informática e software *ad hoc*.³

No caso específico desta pesquisa, o *corpus* a ser utilizado será compilado a partir de um conjunto de produções orais e escritas de textos em espanhol, por parte dos alunos do Curso de Letras/Espanhol, como já mencionado. Após a identificação no *corpus*, os “erros” foram etiquetados conforme uma classificação e categorização. Com o auxílio das ferramentas da LC, as ocorrências de cada tipo de “erro” foram analisadas, com a finalidade de descrever tanto a frequência dos “erros” mais comuns, como os ambientes de ocorrência e detalhes pertinentes à análise gramatical, no sentido de alcançar a interpretação das possíveis causas por sua aparição e/ou persistência.

Com o propósito de tentar encontrar as possíveis causas pela ocorrência de determinados “erros” em detrimento de outros e, ainda, pensando num possível desdobramento para o ensino-aprendizagem de espanhol como língua estrangeira (ELE), o mapeamento de “erros” que a presente pesquisa se propõe realizar poderá constituir-se em importante contribuição. Segundo Durão (2004, p. 19):

O repertório de erros dos alunos não é apenas uma vitrine da insuficiência dos alunos ou as falhas do professor. Pelo contrário, devidamente analisados, os erros são como uma janela aberta sobre o processo de construção de competência na segunda língua; convertido por tudo em uma ferramenta para melhorar os métodos de ensino.⁴

As análises serão desenvolvidas levando em consideração o estudo contrastivo das línguas, materna (LM) e estrangeira (LE), tal como proposto no Modelo de Análise de Erros (CORDER, 1967). A partir da identificação de uma ocorrência de Interlíngua, isto é, identificada por sua idiossincrasia, é realizado um contraste com a forma que seria esperada, tanto em língua materna quanto estrangeira. Desse contraste resultarão as possíveis causas pela ocorrência de “erros”. Para os falantes de português como língua materna, falar uma língua tão

³Minha tradução de “Ahora bien, de partida, afirmo que la LC en su versión actual constituye un enfoque metodológico para el estudio de las lenguas y que presenta oportunidades revolucionarias para la descripción, análisis, y enseñanza de discursos de todo tipo. También brinda una base empírica para el desarrollo de materiales educativos y metodológicos de diversa índole así como para la construcción de gramáticas, diccionarios y otros, tanto de discursos generales como especializados, orales y escritos. Desde esta óptica sostengo que la LC constituye un conjunto o colección de principios metodológicos para estudiar cualquier dominio lingüístico y que se caracteriza por brindar sustento a la investigación de la lengua en uso a partir de *corpus* lingüísticos con sustrato en tecnología computacional y programas informáticos *ad hoc*”. [Observa-se que todas as traduções são de nossa autoria].

⁴“El repertorio de los errores de los aprendices no es solamente un muestrario de la insuficiencia de los estudiantes o de los fracasos del profesor. Muy al contrario, analizados, adecuadamente, los errores vienen a ser como una *ventana* abierta sobre el proceso de construcción de la competencia en la segunda lengua; se convierten por todo ello en un instrumento para mejorar los métodos de enseñanza.”.

próxima como é o espanhol poderá, num primeiro momento, parecer tarefa simples. Segundo Durão (2004, p. 25):

O estudante que se depara com uma língua estrangeira percebe que alguns aspectos da nova língua são muito fáceis, enquanto outros oferecem grande dificuldade. Aquelas características que se assemelham às de sua própria língua, vão parecer fáceis e, por outro lado, aqueles que são diferentes, serão mais difíceis.⁵

A análise contrastiva baseia-se na comparação das LM e LE. Por meio da observação encontram-se as diferenças, as similitudes e as previsões dos possíveis “erros” que são úteis para o ensino-aprendizagem das línguas. Também podem achar interferências linguísticas como produto das similitudes entre a LM e a LE como sugere Baralo (2004, p. 375):

[...] o principal obstáculo para a aquisição de uma língua estrangeira se achava na interferência dos hábitos de língua materna. Por isso, o grau de dificuldade era medido pelo número de estruturas semelhantes entre os dois sistemas de idiomas diferentes. Aquelas que eram iguais na LM e a LE eram transferidos sem problemas; aquelas que fossem diferentes causariam erros por interferência ou transferência negativa. Considerou-se que os erros produziam maus hábitos, por isso era preciso corrigi-los e, sobretudo, evitá-los.⁶

De acordo, então, com a afiliação teórica aos autores anteriormente mencionados e, ainda, a outros referenciais da área que posteriormente serão mencionados no texto da dissertação, pretendemos desenvolver o presente estudo.

A hipótese que norteia esta pesquisa é que o *corpus* de produções orais e escritas de estudantes de ELE oferece o material linguístico necessário para encontrar, a partir da descrição dos “erros” mais frequentes, sua classificação e análise de suas prováveis causas, as evidências de interlíngua.

No caso específico do presente estudo, as primeiras perguntas de pesquisa, cuja resposta procuraremos alcançar por meio de uma abordagem baseada nos princípios da LC, são:

- a) Quais os “erros” mais frequentes na fala e na escrita de aprendizes de ELE?

⁵ “El estudiante que se enfrenta con un idioma extranjero encuentra que algunos aspectos del nuevo idioma son muy fáciles mientras que otros ofrecen gran dificultad. Aquellos rasgos que se parecen a los de su propia lengua le resultarán fáciles mientras que otros ofrecen gran dificultad.”

⁶ “[...] el principal obstáculo para la adquisición de una lengua extranjera se encontraba en la interferencia de los hábitos de la lengua materna. Por ello, se medía el grado de dificultad por la cantidad de estructuras semejantes y de estructuras diferentes entre los dos sistemas lingüísticos. Aquellas que fueran iguales en la LM y en la LE se transferirían sin problemas; las que fueran diferentes ocasionarían los errores por interferencia o transferencia negativa. Se consideraba que los errores producían malos hábitos, por lo que había que corregirlos, y especialmente, evitarlos.”

- b) Quais são as prováveis causas pela ocorrência desses “erros”?
- c) Em que medida uma abordagem de pesquisa baseada em critérios e procedimentos advindos da LC é pertinente para análises de aspectos da Interlíngua de aprendizes de uma língua estrangeira?

Como sugerem Glynn et al (2010, p.19), é preciso definir, pelo menos, perguntas iniciais que poderão ser resolvidas uma vez analisado o *corpus* linguístico:

Isso conduz quatro questões de pesquisa - (i) Que palavras que as pessoas escolhem? (ii) Quais construções gramaticais que as pessoas escolhem? (iii) Como as pessoas usam a palavra? (iv) Como as pessoas usam uma construção gramatical? Podemos rotular estas perguntas de pesquisa, respectivamente, como sinônima lexical, polissemia lexical e gramatical.⁷

Considerando esta breve contextualização da pesquisa, passamos à apresentação dos objetivos, que buscamos alcançar com os trabalhos desenvolvidos. O objetivo mais geral será analisar os “erros” lexicais e gramaticais mais frequentes de aprendizes de ELE e as prováveis causas de sua ocorrência, a partir de um *corpus* oral e escrito composto por produções de universitários do Curso de Letras/Espanhol ILEEL/UFU.

Como objetivos mais específicos destacamos os seguintes:

- a) Identificar as idiossincrasias da interlíngua nas produções orais e escritas dos estudantes;
- b) descrever e analisar as realizações identificadas como interlíngua, num *corpus* de produções orais e escritas de aprendizes, em termos contrastivos com a língua materna e estrangeira;
- c) explicar as prováveis causas pela ocorrência dos “erros” identificados como interlíngua.

O presente texto está composto por uma seção de fundamentação teórica, no qual os seguintes temas serão abordados: Linguística Contrastiva, “Erro” e o Modelo de Análise de Erros, Interlíngua, Teorias de aquisição de línguas estrangeiras e Linguística de *corpus*. A segunda seção descreverá os procedimentos metodológicos e a composição do nosso *corpus* de

⁷ “This leads to four research questions – (i) What words do people choose? (ii) What grammatical constructions do people choose? (iii) How do people use a word? (iv) How do people use a grammatical construction? We can label these research questions respectively as lexical synonymy, grammatical synonymy, lexical polysemy, and grammatical polysemy”.

estudo. Será apresentado em detalhe todo o trabalho metodológico seguido em cada uma das etapas de desenvolvimento do presente trabalho. Na terceira seção será dedicada à análise do *corpus*, com o intuito de ilustrar os caminhos percorridos até a identificação da idiossincrasia da Interlíngua, por meio da análise de erros. Por último, este texto ainda é composto pela presente Introdução, pelas Considerações Finais, as Referências Bibliográficas, Apêndices e Anexos.

2 FUNDAMENTAÇÃO TEÓRICA

O presente trabalho conta com um referencial teórico que passa pelos seguintes pontos específicos, não sendo os únicos que serão desenvolvidos no transcurso do presente ano, mas sim nos quais centraremos a nossa atenção teoricamente. Principalmente baseia-se em temas como a Linguística Contrastiva (aqui representados principalmente em DURÃO, 2004), Interlíngua (BARALO, 2004), Aquisição da língua (KENEDY, 2013; DURÃO, 2004) e Linguística de *Corpus* (PARODI, 2010, 2014).

2.1 Linguística Contrastiva

A Linguística Contrastiva baseia-se na comparação da LM com a LE, ou também entendida como segunda língua (L2). A finalidade dessas comparações é que, dessa maneira, seria possível estabelecer quais são as formas parecidas entre a LM e a LE e quais são as formas diferentes que causariam dificuldade. Segundo Lado (1957 apud Durão, 2004), trata-se de:

Predizer e descobrir quais estruturas causariam dificuldade na aprendizagem e, ao contrário, quais não, por meio do contraste sistemático entre a língua e a cultura objeto de aprendizagem e a língua e a cultura nativas do aluno.⁸

Como afirma Durão (2004), certamente o modelo de Análise Contrastiva (AC) foi desconsiderado como técnica de ensino de segundas línguas, porque não parecia ser um método convincente para os níveis mais avançados de aprendizagem de uma segunda língua; acreditava-se que o método em estágios mais avançados até poderia atrapalhar o processo causando confusão. Contudo, esse método continua vigente, não como uma resposta única para o ensino e aprendizagem de línguas, mas sim como uma base para o desenvolvimento de materiais didáticos para as aulas de língua estrangeira. Llanos (2013), por exemplo, apresenta um guia didático explicativo, ilustrando as inúmeras possibilidades de exploração de um *corpus* oral de espanhol como língua estrangeira, a partir de uma investigação desenvolvida no Laboratório de Linguística Informática da Universidad Autónoma de Madrid, integrando o Modelo de Análise de Erros e a AC.

Segundo Söhrman (2007), a finalidade inicial da Linguística Contrastiva consiste em pesquisar as características das línguas estudadas, para saber quais as diferenças e as

⁸“Predecir y descubrir cuáles estructuras causarían dificultad en el aprendizaje y, al contrario, cuáles no, a través del contraste sistemático entre la lengua y la cultura objeto de aprendizaje y la lengua y la cultura nativas del alumno”.

semelhanças quanto à estrutura, conteúdo e expressão. Para que uma análise contrastiva tenha sucesso, é preciso contar com algumas ferramentas básicas como: boas descrições das línguas (gramáticas extensas); um *corpus* das línguas a serem estudadas compilado criteriosamente; e uma análise dos “erros” que cometem os estudantes. O autor também destaca que é fundamental o bom conhecimento das línguas que serão contrastadas. Não basta somente a análise e descrição de uma língua; para a realização de uma análise contrastiva, é preciso analisar e comparar, pelo menos, duas línguas. Para conhecer o modo como acontece a aprendizagem, é preciso construir um *corpus* de estudantes adequado, isto é, que inclua dados de coleta oral e escrita. Os passos propostos pelo autor, para a realização de uma análise contrastiva são, a saber:

- a) Descrever um determinado fenômeno nas duas línguas estudadas (gramática e léxico);
- b) justapor os fenômenos em ambas as línguas e comprovar a sua equivalência segundo o parecer dos bilíngues;
- c) comparar as línguas em três fases:
 - comparar os sistemas (substantivos, verbos, lexemas específicos, fonemas, etc.);
 - Comparar as construções idênticas (negações, relativos, etc.);
 - Comparar as regras sintáticas idênticas (posição dos adjetivos, etc.).

Conforme destaca Söhrman (2007), nem sempre é possível atender a todos os passos anteriormente citados com o mesmo rigor. Desse modo, é frequente que o pesquisador escolha apenas um dos pontos citados no terceiro item. Com isso, a análise não seria generalista, mas permitiria um aprofundamento maior, pensando, especialmente, em pesquisas de mestrado e doutorado.

Retomando os aspectos pertinentes ao contraste das línguas, é importante saber que existem diversos tipos de semelhanças, entre essas, a saber:

- a) Inteiramente formais (por exemplo, ocorrências diversas de um mesmo som);
- b) homônimas (mesma forma oral) ou homógrafas (mesma escrita, mas com significados diferentes);
- c) semânticas (coincidências e diferenças, em sua maior parte);
- d) funcionais (tem uma função ou objetivos semelhantes).

Uma vez situada a Linguística Contrastiva, passamos à discussão das noções em torno de “*erro*” e as propostas existentes para sua análise.

2.2 Conceitos de “*Erro*” e o Modelo de Análise de Erros

O conceito de “*erro*” é controverso e, com frequência, associado também a aspectos negativos. Por esse motivo, quando a palavra “*erro*” é lida, causa no leitor, às vezes, certo grau de impacto, que poderá levá-lo à estigmatização geral que essa palavra conota. É por isso que se faz necessário esclarecer o significado com que esse conceito será utilizado neste trabalho. Decidimos, então, fazer uma revisão dos conceitos de “*erro*” que anteriormente outros estudiosos utilizaram, no intuito de delimitar conceptualizações nesses trabalhos e de definirmos a nossa afiliação a alguns desses conceitos. Segundo os diversos conceitos de erro que existem, e levando em conta o que Llanos (2013) define como *erro*, não podemos esperar que o processo de aquisição da língua não possua o que esse autor chama de “um ensaio de erro”.

Desse modo, o equívoco se converteria em uma parte essencial para o processo de aprendizagem de uma segunda língua. Llanos (2013, p. 9) também faz uma diferenciação, para podermos entender melhor o que é chamado de “*erro*”, propondo a seguinte classificação: *lapsus*, *falhas* e “*erros*”. Apesar de apresentar três categorias, o autor justifica que, com relação às *falhas* ou *equívocos*, muitos pesquisadores consideram que não há claras diferenças entre *falhas* e “*erros*”, preferindo, portanto, agrupá-los como “*erros*”. Um *lapsus* corresponderia a uma incorreção produzida por motivos extralingüísticos, os quais têm relação direta com o nosso estado mental, no momento de produzirmos um enunciado. Isto é, um *lapsus* pode estar associado, por exemplo, a um estado de cansaço mental ou desconcentração. Por *falhas* ou *equívocos*, Llanos aponta à alternância, num mesmo falante, de formas corretas e incorretas, dentro de um mesmo discurso. Tal fenômeno indica certa insegurança, talvez motivada pela inconsistência na aprendizagem de determinado conteúdo. Com relação ao conceito de *erro*, por outro lado, o autor observa que se trataria de estruturas gramaticais, lexicais ou de formas de pronúncia que não teriam sido adquiridas corretamente. Interpretamos que a noção de *correto*, para esse autor, reside na identificação daquilo que se desvia dos usos reais que fazem os falantes nativos da língua. A diferença entre *lapsus* e *falhas* ou *equívocos* com os “*erros*” é que primeiro se observaria à autocorreção dos próprios falantes.

Segundo Llanos (2013, p. 09), desenvolvendo o conceito de “*erro*”:

Todos erramos, e isso é algo que assumimos com naturalidade. Então, por que um aprendiz de espanhol seria infalível? O processo de aquisição requer um processo de contínuo teste e erro e, quanto mais a língua que se deseja aprender for levada à prática, maior será a probabilidade de se atingir seu domínio... ainda que se continue exposto à possibilidade de cometer mais erros. O erro não deve ser interpretado como algo negativo, porque é uma das caras do processo de aprendizagem.⁹

Uma referência fundamental, bastante citada nos trabalhos revisados na fundamentação teórica do presente estudo, corresponde a Vásquez (1998, 2009a, 2009b). Essa autora (1998, p. 132) contribui também com definições a respeito de “*erros*” e *equívocos*, que ajudam a esclarecer melhor esses conceitos.

Erro: incorreção que se explica pela carência de competência linguística e comunicativa. As regras pertinentes não existem na interlíngua. Nesse sentido toda produção baseada em regras não adquiridas é um erro. *Equívoco*: incorreção que não tem como origem uma falta de competência, mas fatores extralingüísticos. O equívoco se relaciona com a atuação e indica que certas regras adquiridas não têm sido corretamente aplicadas.¹⁰

Comparando os conceitos de ambos os autores, Llanos (2013) e Vásquez (1998), verificamos que aquele considera *lapso* e *fallas* ou *equívocos* o que Vásquez simplesmente chama de *equívocos*. Observando as ressalvas feitas pelos próprios autores e considerando a problema na diferenciação entre *lapso* e *fallas* ou *equívocos*, adotaremos para a presente pesquisa a classificação e o conceito propostos por Vásquez.

A seguir, a Tabela 1 organiza os autores e conceitos anteriormente mencionados:

⁹ “Todos nos equivocamos, y eso es algo que asumimos con naturalidad. ¿Por qué, entonces, un aprendiz de español va a ser infalible? El proceso de adquisición requiere un proceso continuo ensayo y error, y cuanto más se ponga en práctica la lengua que se desea aprender, más probabilidades se tiene de alcanzar su dominio... aun expensas de cometer más errores. Errar no debe interpretarse como algo negativo, puesto que es una de las caras del proceso de aprendizaje”.

¹⁰ “**ERROR**: incorrección que se explica por carencia de competencia lingüística y comunicativa. Las reglas pertinentes no existen en la interlengua. En ese sentido, toda producción basada en reglas no adquiridas es un error. “**EQUIVOCACIÓN**: incorrección que no remite a una falta de competencia lingüística, sino a factores extralingüísticos. La equivocación se relaciona con la actuación e indica que ciertas reglas adquiridas no han sido correctamente aplicadas.” (cf. Corder: error vs. *Mistakes*)

Tabela 1 – Conceitos de lapsos, falhas e erros.

Autor	Lapsos	Falhas	Erros
Vázquez	-	-	Incorreção que se explica pela carência de competência linguística e comunicativa. As regras pertinentes não existem na interlíngua. Toda produção baseada em regras não adquiridas é um erro.
Llanos	Incorreção produzida por motivos extralingüísticos, os quais têm relação direta com o nosso estado mental no momento da produção do enunciado. Geralmente associado a um estado de cansaço mental ou desconcentração.	Se aprecia como uma alternância de formas corretas e incorretas dentro de um mesmo discurso.	Trata-se de estruturas gramaticais, lexicais ou formas de pronunciação que não foram adquiridas corretamente.

Fonte: a autora.

Quanto ao Modelo de Análise de Erros (AE), Corder (1992 [1971]) esclarece os conceitos de Erros Sistemáticos e de Erros não Sistemáticos, constituindo como primeiro tipo de “erros” os lapsos de memória, fadiga física ou mental, que conduzem a pequenos “deslizes” que realmente não constituiriam um reflexo fiel do conhecimento que se possui da língua, mas representaria “um erro casual”. Mas também existiriam os erros sistemáticos, que seriam o ponto fundamental para conhecer um estado de deficiência da língua dos aprendizes.

O modelo de AE concentra sua atenção somente nos “erros” que aparecem na produção dos estudantes. Teve seu ponto de partida em 1967, com os primeiros estudos de Corder. Segundo este autor, os aprendizes de línguas estrangeiras passariam por uma situação parecida à das crianças na hora da aquisição da língua materna, pelo menos no que se refere aos tipos de erros sistêmicos e não sistêmicos que se podem achar. Os critérios da AE que classificam os erros são: linguístico, grammatical, etiológico, pedagógico e comunicativo.

Em Durão (2004), encontramos o modo como o modelo de Análise de Erros propôs as seguintes classificações para os diferentes tipos de *erros* sistemáticos: grammatical, linguístico,

comunicativo, pedagógico e etiológico.

Quanto ao **critério gramatical**, o “erro” afetaria todos os níveis gramaticais, a saber:

- a) Fonológico, quando o “erro” afeta confusões entre fonemas;

(01): *Sí pero la marijuana<FON[marihuana*

Em espanhol, o H não é pronunciado, é mudo. No exemplo, foi pronunciado como fonema fricativo [x], que equivaleria a um J do espanhol “*marijuana*”.

- b) Ortográfico, quando afeta a grafia das palavras;

(02): *A mí no me gusta la idéa<ORT[idea]>*

Em Espanhol, a palavra “idea” não possui acento gráfico. Paroxítonas terminadas em vogal não são acentuadas em espanhol.

- c) Morfológico, quando o “erro” está relacionado com a forma;

(03): *y con esselente<ORT[excelente]> formación*

Trata-se de um “erro” ortográfico, a palavra se escreve corretamente em espanhol como em português, com a combinação “CX”, “excelente”.

- d) Sintático, quando o “erro” está vinculado com a ordem da construção oracional;

(04): *portanto<MOR:SIN[por lo tanto]> tengo*

Em espanhol, esta expressão se escreve separado, “por lo tanto”.

- e) Léxico-semântico, quando o “erro” afeta o significado das palavras.

(05): *Acho <LEX[creo]> que para medicina*

A escolha pela palavra “acho” corresponde a uma transferência do português, Em espanhol, para o caso, seria utilizado “creo” ou “me parece”.

Quanto ao **critério linguístico**, os “erros” podem também corresponder a diversos tipos:

- a) a adição, que acontece quando se adiciona um morfema ou uma palavra redundante ou, simplesmente, o uso da palavra não é adequado ou não faz sentido no contexto no qual está sendo utilizada;
- b) a omissão, que acontece quando um morfema imprescindível no contexto é suprimido;
- c) a ausência de ordem oracional, quando existe uma organização inadequada dos constituintes;
- d) a utilização da forma errada, que indica uma utilização inadequada de um morfema ou palavra num determinado contexto linguístico.

Quanto ao **critério comunicativo**, os “erros” se baseiam no efeito que o enunciado tem no ouvinte; também aqui existem diversas categorias:

- a) global, esse “erro” afeta o enunciado inteiro, tornando incompreensível seu significado;
- b) local, esse “erro” não impede a inferência do sentido da mensagem, mas afeta algum constituinte ou algumas partes do enunciado;
- c) estigmatizante, refere à ideia generalizada que existe acerca de uma determinada comunidade de fala (negros, árabes, judeus, etc.), induzindo a valorações negativas;
- d) irritante, esse “erro” causa enfado porque afeta estruturas do sentido comum;
- e) erro por ambiguidade, gera dificuldade de compreensão devido à existência de ambiguidade;
- f) erro pragmático-cultural, o “erro” é completamente aceitável segundo as normas gramaticais e linguísticas, mas não com respeito à norma pragmática.

Quanto ao **critério pedagógico**, o “erro” pode corresponder a um dos seguintes tipos:

- a) individual, é o tipo de “erro” que caracteriza a interlíngua do falante não nativo. Este “erro” poderia acontecer como uma interferência de línguas aprendidas anteriormente pelo falante;
- b) erro coletivo, esse tipo de “erro” caracteriza a interlíngua de um grupo de estudantes de LE, os quais compartilham a mesma LM. Trata-se de um tipo de erro que poderia ser, inclusive, mais persistente do que o erro *individual*. O erro coletivo se subdivide em dois:
 - o erro coletivo transitório, caracterizado por atravessar vários estágios na aprendizagem da língua, antes de ser superado;
 - o erro coletivo permanente, o qual tem uma tendência a permanecer na interlíngua dos aprendizes. Esse “erro” também pode ser subdividido em dois tipos:
 - i) fossilizável, corresponde à dificuldade em potencial;
 - ii) fossilizado, corresponde à aparição inesperada de formas incorretas que são corretamente aplicadas em certas oportunidades;
- c) erro oral, acontece durante a fala dos estudantes;
- d) erro escrito, que ocorre durante a escrita;
- e) erro de compreensão, esse tipo de “erro” origina-se na recepção da mensagem, basicamente corresponde a uma decodificação errada do que foi ouvido ou lido;
- f) erro de produção, acontece quando os aprendizes propõem enunciados total ou parcialmente errados.

Com relação ao **critério etiológico**, os “erros” podem ser classificados da seguinte forma:

- a) Transferência/interferência/intralingüístico: os “erros” podem ocorrer das seguintes maneiras:
 - a similitude fonológica ou ortográfica acontece, quando uma palavra da LE é semelhante a uma palavra da LM, em sua forma oral ou escrita;
 - a extensão por analogia acontece quando, no momento de propor ou interpretar uma determinada forma, tem-se em mente a forma da LM;
 - a falta de habilidade para distinguir aspectos gramaticais da LE com respeito aos da LM;

- a utilização de estrangeirismo nem sempre corresponde a uma forma errada, mas é considerado errado, quando usa-se um estrangeirismo por desconhecimento da forma correta na língua objeto;
- b) Intralinguístico/intralingual: podem acontecer pelos seguintes motivos:
 - por simplificação, ocorre devido a uma aplicação das regras de forma indevida ou falta de aplicação das regras;
 - por generalização (ou hipergeneralização), extensão indevida das regras, nos casos de exceção;
 - por indução, acontece numa orientação errada ou parcial, oferecida por um professor ou por um material didático;
 - por produção excessiva, trata-se de um “erro” estilístico, consiste na repetição exagerada de uma forma determinada.

Para o presente trabalho de dissertação, consideraremos o seguinte critério, dentre os mencionados acima: o critério gramatical, será o único e principal, pois o *corpus* será etiquetado de acordo as categorias que aqui se propõem. Os outros critérios serviram para dar um panorama geral dos tipos de erros que existem, mas para delimitar a análise, somente será utilizado o critério gramatical.

Corder (1967 apud BARALO, 2004) destaca a importância dos “erros” tanto para os alunos, como para os professores e os investigadores acerca do processo de aprendizagem de uma LE, pois percebia uma espécie de dialeto idiossincrásico, com suas próprias peculiaridades, diferentes da LM e da língua objeto de estudo. Baralo (2004) aponta que, para esse pesquisador (que se inspirou nos trabalhos de Chomsky), a primeira etapa da AE é o reconhecimento da idiossincrasia. A segunda é a descritiva, momento em que se compara o dialeto idiossincrásico do aluno comparando-o com sua LM e com a LE. A terceira etapa seria aquela em que se atinge o objetivo da AE, explicativa, quando se deveriam encontrar os fundamentos psicolinguísticos do como e do porquê do dialeto idiossincrásico.

Mesmo que a proposta teórica de Corder (1992 [1967,1971, 1976]) seja uma referência que já possui praticamente meio século de existência, ainda representa um referencial importante nos estudos que abordam a aquisição de segundas línguas e, principalmente, as abordagens que consideram a análise de “erros”. Prova disso são os múltiplos estudos que ainda tomam por referência esse autor. São diversas as pesquisas baseadas no modelo de análise proposto por Corder. Alguns desses trabalhos, em que o autor é citado como referência

obrigatória, são teses doutorais, livros e artigos, consideravelmente recentes, a saber: Vásquez (1998, 2009), Baralo (1999, 2004), Yokota (2001), Durão (2004, 2007), Andrade (2011), Rufat (2015), Camarena (2011), Llanos (2013).

A próxima subseção busca reunir conceitos de *Interlíngua* e noções características que definiram o estabelecimento dessa área.

2.3 Interlíngua

Uma vez definidas as noções de “*erro*” e conhecido o modelo proposto para sua análise, cabe reconhecer as condições para sua produção. Isto é, para completar a compreensão do significado que engloba aquilo que se denomina “*erro*”, é necessário considerar o contexto em que esses “*erros*” seriam produzidos. Segundo Selinker (1992), a ocorrência dos “*erros*” aconteceria em uma estrutura psicológica latente, potencial de aparecer em algum momento da vida de todos os adultos que tentam aprender segundas línguas. Tal estrutura não é regida pelas regras da Língua Materna ou da Língua Objeto. Trata-se de uma estrutura diferente, latente, em que começa a se configurar a aprendizagem de uma nova língua com suas regras próprias. Essa estrutura psicológica latente comportaria a chamada *Interlíngua*.

Söhrman (2007, p. 17), explica o que entende por interlíngua:

Outro fenômeno de envergadura importantíssima é a interlíngua, que é o reflexo da sistematização linguística na competência do aprendiz no momento determinado em que faz generalizações e comete erros sintáticos. Ao mesmo tempo a interlíngua é uma norma pouco estável, a qual vai mudando quando o estudante avança até o domínio da língua estudada.¹¹

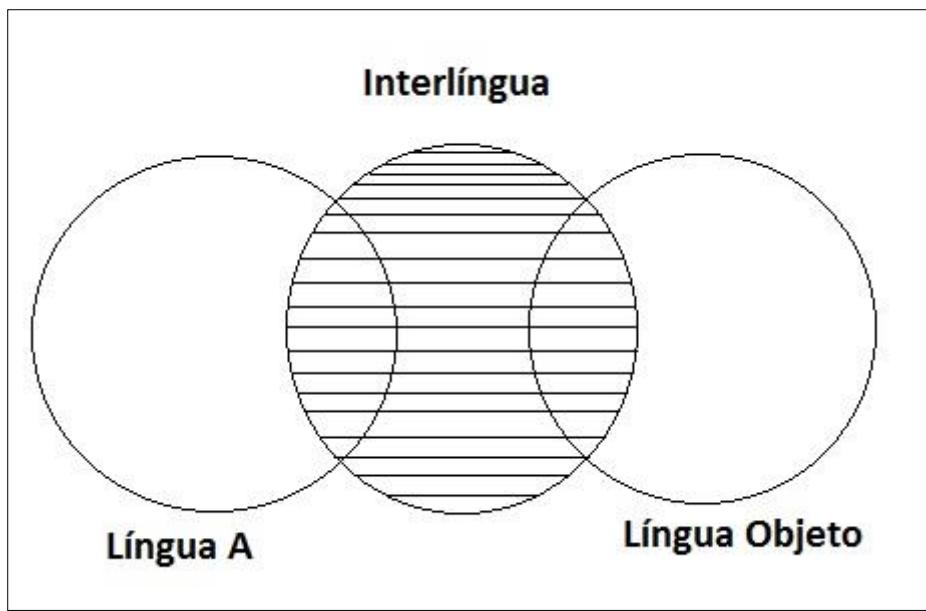
Corder (1992 [1971]) emprega uma nomenclatura diferente para esse fenômeno. O autor prefere a denominação de *dialecto idiossincrático*, que corresponderia a um sistema com gramática e regras assim como ocorre com outras gramáticas, mas com regras próprias que não se acham em dialetos padrões. Podem ser considerados *dialetos idiossincráticos*, por exemplo: a linguagem poética, a linguagem dos afásicos, a linguagem das crianças, assim como a linguagem das pessoas que aprendem segundas línguas. No caso das crianças e também dos aprendizes de segundas línguas, pode-se falar em *dialecto transicional* porque, diferente do caso da linguagem poética, os enunciados idiossincrásicos não podem ser considerados como

¹¹ "Otro fenómeno de envergadura importantísima es la interlengua que es el reflejo de la sistematización lingüística en la competencia del aprendiz en el momento determinado en que hace generalizaciones y comete errores sintácticos. Al mismo tiempo la interlengua es una norma poco estable, que va cambiando cuando el estudiante avanza hacia el dominio de la segunda lengua estudiada"

voluntários, nem tampouco como um desvio linguístico, como é o caso das pessoas afásicas. Os “erros” são considerados, portanto, como parte do processo de aquisição. Desse modo, os “erros” constituiriam como uma prova de que, no processo de aprendizagem, os hábitos corretos ainda não teriam sido adquiridos.

A Figura 1, a seguir, ilustra o espaço que ocuparia a Interlíngua:

Figura 1 – Interlíngua



Fonte: Corder, 1992, p. 68 [1971].

O dialeto idiossincrático transicional dos aprendizes de segundas línguas, para Corder (1992 [1971]), é caracterizado por ser regular, sistemático, por possuir significado, ter uma gramática e por possibilitar potencialmente a sua descrição. Já Selinker (1992 [1969]), por sua vez, propôs dar a esse tipo de dialeto idiossincrático o nome de *interlíngua*, considerando, com isso, que se trata de um dialeto e que suas regras compartilham características com outros dois dialetos sociais ou com duas línguas.

Segundo Durão (2004), falar em interlíngua significa fazer referência a um estado da língua estrangeira que está sendo aprendida, que pode ser compreendido também como um conceito construído depois da aparição dos conceitos de AC e AE, respectivamente; isto é, corresponde a mais um estágio no processo de aquisição da língua meta. “A interlíngua de aprendizes de línguas manifesta construtos coerentes de normas, que em cada momento do processo de ensino-aprendizagem são sistemáticos e, ao mesmo tempo, variáveis” (DURÃO,

2004, p. 61). ¹²

Segundo Nemser (1971 apud DURÃO, 2004), a transferência é um fenômeno muito frequente quando se aprende uma nova língua, acontece de maneira natural, segundo o fundamento teórico do AC. A explicação desse fenômeno consiste em que os conhecimentos adquiridos anteriormente têm influência direta no novo conhecimento que está sendo aprendido, podendo tratar-se de dois tipos de transferência: positiva e negativa. A transferência pode ser de tipo positiva quando o conhecimento prévio coincide com a nova forma que está sendo aprendida e, portanto, pode ser proveitoso para o novo conhecimento a ser aprendido. Ou, pode tratar-se de uma transferência de tipo negativo, quando o conhecimento aprendido anteriormente é oposto ao novo e, em decorrência disso, dificulta mais a aquisição e leva a cometer novos “erros”.

É possível encontrar na arte várias pesquisas que se assemelham ao presente estudo com o qual me proponho finalizar a minha dissertação de mestrado. Consequentemente, foi possível realizar pontos de comparação com a minha própria pesquisa, tomar bons exemplos e, sobretudo, considerá-las também dentro do marco teórico como, por exemplo, um estudo realizado no México por Camarena, em 2011. Trata-se de um estudo baseado nos “erros” de interlíngua produzidos nas produções escritas feitas por estudantes universitários durante o processo de aprendizagem de português, com o intuito de estabelecer as possíveis causas e poder resolver diferentes problemas de docência. Seu valor principal é ser um referente para saber quais são os problemas principais na aula de idiomas e que, portanto, devessem ser revisados com mais detalhes por parte do professor.

2.4 Linguística de *Corpus*

Outro dos pressupostos teórico-metodológicos importante que faz parte desta dissertação é a Linguística de *Corpus*. Ainda que não se trate exatamente de uma teoria linguística, embora muitos pesquisadores assim o entendam, principalmente por possibilitar um olhar sobre a língua em termos probabilísticos, a LC é assumida como uma metodologia e/ou abordagem. Um dos seus principais atributos seria a capacidade de subsidiar praticamente qualquer uma das áreas da linguística, isto é, a LC é muito ampla na sua aplicabilidade, dentro da grande área da linguística.

¹² “La interlengua de aprendices de lenguas manifiesta constructos coherentes de reglas, que en cada momento del proceso de enseñanza/aprendizaje son sistemáticos y, al mismo tiempo, variables”.

Em Berber Sardinha (2004), o autor faz uma resenha histórica acerca da LC, que nos leva a refletir sobre alguns fatos de importância como, por exemplo, que a LC já existiria há muito tempo, prova disso são os destacados linguistas como Boas e Fries. Mas a grande diferença entre a LC desses tempos com a da atualidade consiste, basicamente, em que agora contamos com a ferramenta tecnológica que facilita a nossa coleta e armazenamento de *corpora*, o computador.

Para Berber Sardinha (2004, p. 3),

A Linguística de *Corpus* ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador.

Outra diferença da versão anterior da LC com a atual reside na finalidade que os *corpora* tinham. Anteriormente eram utilizados principalmente com fins pedagógicos, mais especificamente voltados para o ensino de línguas estrangeiras. Hoje em dia, a finalidade passou a ser principalmente a descrição de línguas, mas também pode-se mencionar um ressurgimento da LC com fins pedagógicos, talvez com menos ênfase do que antigamente.

Um panorama da LC também foi feito pela autora Shepherd (2012), apontando que a LC desde seus origens foi associada a programas de computador que processavam grandes textos ou compilações de textos pequenos e aos linguistas que deviam ter habilidade para ler linhas de concordância. Os objetivos de estudo baseavam-se inicialmente na lexicografia, mas hoje em dia abrange múltiplas áreas da linguística.

Também faz menção sobre a representatividade dos *corpora*, pois, segundo ela, não é mais definida simplesmente pelo seu tamanho, pois existem interessantes estudos feitos com *corpora* pequenos, como exemplo, os *corpora* de aprendizes ou textos literários de um determinado autor.

Na última década adquire força o estudo das unidades “multilexicais”, identificando quais são ou seriam grupos de palavras com significado com os quais se pode trabalhar. Desde que se começou a analisar compostos lexicais nas análises de *corpora*, os linguistas saíram da sua zona de conforto.

Quanto às gramáticas, passam ao largo da exemplificação da linguagem em uso, como estratégia para explicar algum fenômeno linguístico. Com as *corpora* de e para aprendizes, foram-se criando e desenvolvendo novos materiais didáticos, agora preferenciando os sentidos mais frequentes sobre as palavras mais frequentes.

A confecção destes materiais passaram por diversas etapas, desde os poucos motivadores dicionários pedagógicos, até incluir os *corpora* de aprendizes como base de materiais de aprendizagem de línguas estrangeiras. Mas além do uso que se dá à LC para os fins de criação de material didático, a LC também é utilizada em estudos de discurso a partir de léxico determinado, em tradução, com uma utilização de *corpora* que permite fazer análise de estilo, entre outros.

A autora encerra seu panorama da linguística fazendo referência a que o linguista de *corpus* deve possuir uma habilidade analítica desenvolvida e, além disso, conhecimento de sistemas operacionais básicos, principalmente relacionados aos softwares de *corpora*.

Em entrevista concedida no âmbito do XII Encontro de Linguística de *Corpus* (ELC) e a VII Escola Brasileira de Linguística Computacional (EBRALC), realizados no ILEEL/UFU em novembro de 2014, Parodi apresenta conceitos essenciais caros à LC:

É uma coleção de documentos que deveria estar alinhada com alguns princípios relacionados com a sua própria compilação, de maneira que não deveria ter uma heterogeneidade no seu caráter e sim alguns princípios unificadores nesse sentido, e deveria, por exemplo, responder a algumas perguntas – como já afirmamos – que o pesquisador compilador do *corpus* precisaria fazer a si mesmo. O ideal seria ter a maior quantidade de informação possível, para poder saber o que é, de fato, um *corpus*, ou qual é o tipo de *corpus* que está sendo compilado. Também deveríamos ter alguns parâmetros ecológicos, ou seja, o ideal é que os textos da variedade que estamos compilando e que conformam o *corpus* sejam completos. Isto é, se são conversações de algum tipo de registro oral, tomara que sejam conversações completas, ou seja, que de acordo com os propósitos comunicativos que conduzem o diálogo, possamos contar, do princípio até o fim dessa conversação, com todos os participantes, quantos eles forem, e estejam registrados. (NOVODVORSKI; HERRERA, 2014, p. 457-458)¹³.

Na mesma entrevista, o linguista também aproveitou para registrar o seu próprio entendimento por *Linguística de Corpus*.

Em minha opinión, a LC é uma metodología muito potente, para mí a LC não es una teoría. Es una forma de trabajar, no tiene necesariamente un conjunto de principios teóricos. De ese modo, posso trabajar con la LC a partir

¹³ “Es una colección de documentos que debe estar alineada con algunos principios que tienen que ver con la recolección misma de este, de modo que no debiera tener una heterogeneidad en su carácter, sino que debiera tener algunos principios unificadores en este sentido y debiera, por ejemplo, responder a algunas preguntas como ya decíamos que tiene que hacerse el investigador que lo recolecta. Ojalá tuviéramos la mayor cantidad de información posible para poder saber qué es un *corpus*, o qué *corpus* es el que estamos juntando; también debiéramos tener algunos parámetros ecológicos, o sea, ojalá que los textos que conforman el *corpus* sean textos completos de la variedad que estamos recolectando, o sea, si son conversaciones de algún tipo de registro oral, ojalá sean conversaciones completas, o sea, que de acuerdo a los propósitos comunicativos que la guían, tengamos, desde el principio al fin de esa conversación, donde los participantes, cuantos sean, sean los que están registrados.”

da linguística sistêmico-funcional, da teoria do gênero, ou desde outras perspectivas linguísticas. Portanto, quando tomamos essa decisão, surge claramente a ideia de que a LC não é um paradigma linguístico, no sentido kuhniano, ou seja, que tenha um objetivo, um objeto e um método específico, e que tenha uns princípios teóricos que a orientem. Acredito que é, acima de tudo, uma forma de se aproximar ao tratamento da informação que é coletada, e isso faz com que seja mais poderosa ainda, porque pode atravessar todas as áreas de pesquisa da linguística [...].¹⁴

Verificamos, na citação anterior, que Parodi (2014) discute o lugar que ocuparia a LC, assumindo-a como uma abordagem, “uma forma de se aproximar ao tratamento da informação que é coletada”. O pesquisador ainda enfatiza o poder da LC, uma vez que sua condição lhe permitiria “atravessar todas as áreas de pesquisa da linguística”. Tal afirmação é altamente relevante, porque destaca um papel de importância para a LC, na interface com as diversas áreas e correntes da linguística em geral.

Um exemplo desse diálogo entre a LC e outras áreas é Fernández (2011), que descreve a perspectiva de análise da aquisição de língua materna com subsídios da LC. A autora apresenta a importância dos *corpora* orais, a partir do seu estudo do *Corpus Koiné* de fala infantil. Nele, a autora manifesta que a LC vem-se focando na observação de situações comunicativas reais, sendo o seu objetivo os dados verbais concretos e, em decorrência disso, a necessária inovação nos diversos métodos e enfoques. Uma consequência é que, com o tempo, os procedimentos focados apenas na linguagem escrita, considerada como um produto estático, vêm tornando-se obsoletos. A necessidade de consulta a *corpora* orais vem afirmando-se cada vez mais no cenário de pesquisa.

Por esse motivo, a pesquisadora destaca que a linguística dos usos, uma vez que tem como interesse central a observação da fala real e os seus componentes, alcança uma maior efetividade na análise de aspectos comunicativos. Dessa maneira, não há lugar para os preconceitos acerca de emissões não acordes com um determinado padrão prefixado; todo intercambio verbal, com substrato comunicativo, passa a ter importância.

Quanto a ferramenta da LC que se utilizou para o presente trabalho, Sardinha (2006) realizou uma abordagem sobre o programa utilizado para o processamento dos dados. Ele

¹⁴ “En mi opinión, la LC es una metodología muy potente, para mí la LC no es una teoría. Es una forma de trabajar, no tiene necesariamente un conjunto de principios teóricos, de modo que yo puedo hacer LC desde la lingüística sistemática funcional, desde la teoría del género, o desde otras perspectivas lingüísticas. Por tanto, cuando lo decimos así, emerge claramente la idea que la LC no es un paradigma lingüístico en el sentido kuhniano, o sea que tenga un objetivo, un objeto y un método específico, y que tenga unos principios teóricos que la orientan; creo que es más bien una forma de aproximarse al tratamiento de la información que se recolecta, y eso hace que sea más poderosa aun, porque puede estar transida en todas las áreas de investigación lingüística [...].”.

apresenta o programa *WordSmith Tools (WST)* como um conjunto de programas, em um programa só, ou seja uma “suíte”, o qual tem como objetivo facilitar a análise linguística baseada na frequência e na co-ocorrência das palavras. Mas além disto, também possui outras funcionalidades como permitir ajustar o *corpus*, tirar e adicionar elementos, colocar e remover etiquetas, entre outros.

WST é um software que possui um grande número de usuários e que no Brasil se fez popular com múltiplos cursos, oficinas e palestras ministradas com a finalidade de ensinar a utilização do programa.

Wordlist, Keyword e Concord são as ferramentas principais do WST. Retomando a Sardinha (2006, p. 9):

WordList: produz listas de palavras contendo todas as palavras do arquivo ou arquivos selecionados, elencadas em conjunto com suas frequências absolutas ou percentuais. Também compara listas, criando listas de consistência, onde é informado em listas cada palavra aparece.

Concord: realiza concordâncias, ou listagens de uma palavra específica quantas (o “nódulo”, *node word* ou *search word*) juntamente com parte do texto onde ocorreu. Oferece também listas de colocados, isto é, palavras que ocorreram perto do nódulo.

KeyWords: extrai palavras de uma lista cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras num outro *corpus* (de referência). Calcula também palavras-chave, que são chave em vários textos.

O WordSmith Tools é somente uma ferramenta computacional que ajudaria a análise, mas jamais a esgota com o simples processamento dos dados que ele efetua.

A próxima seção traz uma descrição detalhada dos procedimentos metodológicos desenvolvidos, desde o levantamento e leitura do referencial teórico, até a finalização da escrita do presente trabalho de dissertação, para a compilação e análise do *corpus* de estudo.

3 METODOLOGIA E *CORPUS*

Com o intuito de organizar essa seção e deixar esclarecido quais foram cada um dos passos na presente pesquisa, desde o envio do projeto ao Comitê de Ética em Pesquisa (CEP) até a redação da presente pesquisa. A seguir, pontuam-se cada uma das fases:

Procedimentos metodológicos gerais:

- a) levantamento e leitura do referencial teórico;
- b) envio do projeto para aprovação no Comitê de Ética em Pesquisa;
- c) planejamento das atividades para a coleta de dados;
- d) aplicação do Questionário de Sondagem;
- e) composição do grupo de estudantes que participaram da pesquisa;
- f) assinatura do Termo de Consentimento Livre e Esclarecido (TCLE), por parte dos participantes envolvidos na pesquisa;
- g) gravação das produções orais;
- h) produção dos textos escritos;
- i) definição de critérios para armazenamento do *corpus*;
- j) definição de critérios e códigos para a transcrição das produções orais;
- k) transcrição ortográfica dos arquivos sonoros;
- l) digitação dos textos escritos;
- m) etiquetagem do *corpus* de estudo;
- n) eliminação dos arquivos sonoros com as produções orais dos participantes;
- o) devolução dos textos escritos aos participantes da pesquisa;
- p) processamento dos arquivos transcritos e digitalizados com as ferramentas da Linguística de *Corpus*;
- q) levantamento dos dados;
- r) identificação das ocorrências segundo os critérios escolhidos para as análises;
- s) análise de (sub)categorias e classificação;
- t) levantamento das prováveis causas na ocorrência da interlíngua de aprendizes de ELE;
- u) redação de artigos e apresentação em eventos para divulgação da pesquisa;
- v) redação da dissertação.

A natureza empírica deste estudo, como já apontado, assume um prisma analítico-descritivo, com subsídios da Linguística de *Corpus* e, mais especificamente, adotando uma perspectiva de *Corpus-Driven*, isto é, guiados pelo *corpus*. Para Glynn et al (2010, p. 9):

Os métodos quantitativos são essenciais para determinar a probabilidade de que nossas generalizações são representativas da população, mas também porque ajudam a encontrar padrões de co-ocorrência que poderiam ser de outro modo impossíveis de identificar.¹⁵

Segundo Tagnin (2011, p. 358-359), o estudo baseado em *corpus* (*corpus based*) é aquele em que “o *corpus* é usado (a) para comprovar (ou não) uma hipótese ou (b) para extrair exemplos”; já os estudos direcionados ou guiados por *corpus* (*corpus driven*) são aqueles em que “o estudo se desenvolve conforme dados apresentados pelo *corpus*, sem pressuposições teóricas”.

Nosso primeiro procedimento metodológico consistiu no levantamento e leitura do referencial teórico. Para tanto, delimitou-se que, para o presente estudo, a fundamentação teórica que daria principal sustento às análises descritivas seria: Interlíngua (CORDER, 1967; SELINKER, 1972; BARALO, 1999, 2004; DURÃO, 2007), Linguística Contrastiva (SÖHRMAN, 2007), Modelo para Análise de Erros (DURÃO, 2004; ANDRADE, 2011; SANTOS GARGALLO, 2004), dentre os principais. Adotamos uma perspectiva de análise de base empírica, apoiados nos subsídios que propicia a Linguística de *Corpus* (BERBER SARDINHA, 2004; PARODI, 2010), entendida aqui como metodologia e abordagem para a pesquisa.

Depois de realizar o levantamento e iniciarmos as leituras e discussões em torno do referencial teórico, submetemos o projeto de pesquisa ao CEP para sua aprovação. Enquanto aguardávamos a aprovação do projeto por parte do CEP, iniciamos o planejamento das atividades que seriam realizadas após a dita aprovação.

Planejamos sete atividades no total, quatro de produção oral e três de produção escrita. Quanto às atividades de produção oral, tivemos de contar com um laboratório de idiomas que nos permitisse fazer gravações de áudio simultâneas, de maneira que fosse possível gravar por separado a cada um dos participantes, assim como também efetuar gravações em coletivo. A primeira atividade desenvolvida foi denominada “apresentação”, nela

¹⁵ “Quantitative methods are essential in determining the probability that our generalizations are representative of the population but also because they help find patterns of co-occurrence that might be otherwise impossible to identify”.

as gravações foram feitas de maneira simultânea no Laboratório de Línguas (LAMILING), mas separadas as gravações de cada um dos participantes. Na atividade, pediu-se para os participantes que, durante um tempo máximo de três minutos, fizessem uma apresentação deles próprios, sugerindo que incluíssem detalhes, por exemplo, quais eram suas ocupações, quais os passatempos, falar sobre a família, entre outros.

A segunda atividade de produção oral proposta aos participantes também foi gravada em simultâneo, mas separadas as gravações de cada um deles. A atividade foi denominada “leitura”, nela foi apresentado um mesmo texto para cada um dos participantes, contendo a informação sobre a aprovação do uso da maconha no Uruguai e a regulamentação dela. Pediu-se para os participantes lerem o texto de maneira tranquila, pois a atividade não teria um tempo máximo para a gravação.

A terceira atividade foi realizada em duplas, denominada “conversação”. Nela, pediu-se para os participantes que simulassem a seguinte situação: um professor de espanhol chega para trabalhar em uma escola onde a disciplina de espanhol como língua estrangeira é estudada pela primeira vez no estabelecimento educacional, a recepção para o novo professor devia ser feita pelo professor de história do estabelecimento que lhe explica como é o funcionamento da escola. O tempo determinado para essa atividade era de, no máximo, seis minutos.

A quarta e última atividade de produção oral teve um caráter coletivo, no qual todos os participantes do período deviam interatuar em um debate. Para essa atividade, denominada “debate”, pediu-se para os participantes assistirem um vídeo de uma entrevista com o ex-Presidente do Uruguai, José Mujica, que explicava o porquê da aprovação da lei que permitiria o uso recreativo da maconha. Além disso, pediu-se que considerassem o texto já lido na atividade de “leitura” sobre a regulamentação do uso da maconha. Separou-se o grupo de participantes entre quem apoiaria a liberação do uso da maconha no Brasil com os que não a aprovariam. Dessa forma, iniciou-se um debate que teve um tempo máximo de duração de oito minutos.

Quanto às atividades de produção escrita, foram aplicadas imediatamente após concluir a sessão de produção oral, ou seja, utilizando o mesmo espaço do LAMILING. Para tais, utilizamos o formato papel como suporte do instrumento a ser aplicado.

Para este propósito planejamos três atividades, sendo a primeira nomeada “opinião”. Nela os participantes deviam ler um texto de meia página que tratava da mercantilização da variante oficial da língua espanhola. Feita a leitura do texto, eles deviam

responder duas perguntas de opinião à raiz do texto, em um máximo de doze linhas.

A segunda atividade de produção escrita foi denominada “descrição”. Nela, solicitou-se ao participante fazer uma descrição da UFU e do Curso de Letras. Assim como na primeira, os participantes tiveram um máximo de doze linhas para responderem.

A terceira atividade de produção escrita foi denominada “carta”. Nela, foi solicitado aos participantes para se colocarem na situação hipotética de estar em algum país da América Latina trabalhando como professores de português e escreverem uma carta contando as suas impressões do lugar e do trabalho. O espaço máximo para desenvolver a atividade foi de vinte linhas.

Uma vez aprovado o projeto no CEP (ver ANEXO com parecer consubstanciado) e obedecendo aos critérios estabelecidos no projeto, foram realizadas visitas a cada um dos períodos que iriam fazer parte deste trabalho (segundo, quarto, sexto e oitavo períodos), para apresentação da pesquisa, explicando os seus objetivos e solicitando/aguardando a colaboração espontânea no projeto. Na ocasião, aplicamos o Questionário de Sondagem (ver APÊNDICE), que tinha como objetivo efetuar uma seleção dos participantes que obedecessem, de modo geral, aos seguintes critérios: os informantes deveriam ser estudantes do Curso de Letras/Espanhol do ILEEL/UFU, durante a época de realização da presente pesquisa, regularmente matriculados.

O critério para seleção foi a manifestação espontânea de interesse em participar da pesquisa e que os estudantes não tivessem feito um estudo regular da língua espanhola antes de ingressar no Curso de Letras. O intuito dessa limitação, quanto à ausência de estudos prévios em língua espanhola, foi o de equilibrar o tempo de aprendizagem da língua espanhola em aulas formais. Desse modo, os alunos de cada período teriam cursado, em tese, o mesmo tanto de horas de espanhol. Uma vez analisados os questionários, reunimos o grupo de estudantes que participariam da pesquisa.

Em relação às horas de aulas formais de língua espanhola no Curso de Letras Habilitação em Espanhol, há um incremento na carga horária contínuo no decorrer do Curso de Letras. Podemos apreciar que, no primeiro período, há apenas duas disciplinas da especialidade, a primeira “Língua Espanhola: Aprendizagem Crítico-Reflexiva” com sessenta horas teóricas na carga horária, com ementa descrita da seguinte forma: “Esta disciplina proporciona o debate sobre questões culturais, políticas, sociais e cognitivas relacionadas com o processo de aprendizagem de língua espanhola, de forma que os alunos possam despertar para suas expectativas e reais necessidades de aprendizado da língua alvo, refletindo, ainda,

sobre suas dificuldades e possibilidades de uso de Espanhol (falado e escrito) quando se vive em um país no qual essa língua parece não fazer parte do seu cotidiano. A abordagem adotada contará com o uso de textos autênticos (em língua materna e língua Espanhola) que tratam das questões propostas.” e; a segunda disciplina “Pipe 1: Línguas estrangeiras”, consta de trinta horas práticas na carga horária, descrita, na ementa, da seguinte maneira: “Compreensão dos conceitos de Anglofonia, Francofonia e Hispanofonia. Estudo de aspectos históricos, geográficos, políticos, econômicos, sociais, linguísticos e culturais de países falantes das línguas Inglesa, Francesa e Espanhola.”.

No segundo período também encontramos duas disciplinas da especialidade, uma de caráter teórico e a outra de caráter prático, tratando-se então de “Língua Espanhola: Fundamentos Linguísticos”, a qual consta de sessenta horas teóricas na carga horaria e descrita, na ementa, da maneira que segue: “Estudo dos mecanismos de funcionamento da Língua Espanhola, dentro de uma perspectiva que considere a língua como instrumento de comunicação e como uma prática social constituída também por uma dimensão ético-política.” e; da disciplina nomeada “Pipe 2: Línguas Estrangeiras” contando com trinta horas de caráter prático na carga horária, descrita, na ementa, da seguinte maneira: “Condições de trabalho do professor de língua estrangeira em escolas ou institutos de idiomas. O papel do professor de língua em escolas ou institutos de idiomas no contexto educacional brasileiro. Concepções de ensino e de aprendizagem de línguas estrangeiras.”. Desta forma, no final do primeiro ano, totalizam-se 180 horas teóricas e práticas da especialidade.

No segundo ano de estudos do curso, no terceiro período, pode-se apreciar uma diminuição na carga horária, existindo apenas duas disciplinas da área, a primeira com sessenta horas teóricas na carga horaria e a segunda com apenas quinze horas práticas na carga horária, respetivamente: “Língua Espanhola: Habilidades Integradas Com Ênfase na Compreensão Oral”, e a nomeada “Pipe 3: Fonética e Fonologia da Língua Espanhola”. A ementa da primeira é descrita como: “Esta disciplina tem seu foco predominante no desenvolvimento da habilidade de compreensão oral em Língua Estrangeira, embora as demais habilidades não sejam negligenciadas. São trabalhados os conhecimentos: estratégico, textual, contextual, atitudinal e sistêmico, de forma a possibilitar a inserção e participação comunicativa do aluno em relação a diferentes gêneros, nas suas dimensões textuais e discursivas, que permeiam sua vida pessoal, profissional e acadêmica, tanto em ambientes presenciais como aqueles mediados pelas novas tecnologias.”; já a segunda: “Conhecimentos de fonética (acústica e articulatória) e fonologia da língua espanhola e utilização do Alfabeto Fonético Internacional em leitura de transcrições

fonéticas de textos diversos. Verificação da importância dada ao estudo da fonética e fonologia em Livros Didáticos e materiais gravados em áudio”.

Já no quarto período, aprecia-se um aumento na carga horária. O aluno deve cursar três disciplinas específicas da habilitação em espanhol, as quais possuem sessenta horas teóricas cada, sendo: “Língua Espanhola: Habilidades Integradas Com Ênfase na Produção Oral” a qual está descrita, na ementa, da seguinte maneira: “Foco predominante nas habilidades e estratégias para comunicação oral em Língua Espanhola, embora as demais habilidades não sejam negligenciadas. Trabalho pedagógico para desenvolvimento e ampliação de conhecimentos estratégicos, textual, de mundo, atitudinal e sistêmico, de forma a otimizar e possibilitar a participação comunicativa dos graduandos na produção de gêneros discursivos que permeiam sua vida pessoal, profissional e acadêmica.”; pensando no desenvolvimento das habilidades de leitura, apresenta-se a disciplina “Língua Espanhola: Habilidades Integradas Com Ênfase na Leitura”, descrita na ementa como: “Esta disciplina tem seu foco predominante no desenvolvimento da habilidades de leitura em Língua Espanhola, embora as demais habilidades não sejam dispensadas. São trabalhados os conhecimentos estratégicos pertinentes ao processo de leitura, textual, de mundo, atitudinal e sistêmico, de forma a possibilitar a inserção e participação comunicativa dos alunos em relação aos gêneros discursivos, nas suas dimensões textual e discursiva, que permeiam sua vida pessoal, profissional e acadêmica tanto em contextos presenciais como nos mediados pelas novas tecnologias.” e; uma disciplina focada em desenvolver as habilidades na escrita da língua estrangeira, “Língua Espanhola: Habilidades Integradas Com Ênfase na Escrita”, descrita na ementa da seguinte forma: “Esta disciplina tem seu foco predominante no desenvolvimento da habilidade de escrita em língua estrangeira, embora as demais habilidades não sejam dispensadas. São trabalhados os conhecimentos estratégicos de produção escrita, textual, de mundo, atitudinal e sistêmico, de forma a possibilitar a inserção e participação comunicativa dos alunos em relação aos gêneros discursivos, em suas dimensões textual e discursiva, que permeiam sua vida pessoal, profissional e acadêmica, tanto em ambientes presenciais tanto em ambientes mediados pelas novas tecnologias.” Desta forma, no final do segundo ano de estudos, totaliza-se 255 horas exclusivas da especialidade.

No quinto período pode-se apreciar um incremento ainda maior na carga horária da especialidade de espanhol. Com um currículo quase exclusivamente focado na especialidade da língua espanhola, os alunos do quinto período devem cursar as seguintes disciplinas: “Língua Espanhola: Práticas Discursivas do Cotidiano” a qual contém um total de sessenta

horas teóricas na carga horaria, definida na ementa da seguinte maneira: “Nesta disciplina as habilidades de leitura, escrita, produção oral e compreensão oral serão trabalhadas tendo em vista as práticas discursivas do cotidiano em Língua Espanhola que circulam nos contextos presenciais e nos contextos mediados pelas novas tecnologias”; com sessenta horas teóricas na carga horária, a disciplina “Língua Espanhola: Práticas Discursivas da Academia”, com a seguinte ementa: “Nesta disciplina as habilidades de leitura, escrita, oralidade e compreensão oral serão trabalhadas em termos das práticas discursivas do contexto acadêmico em língua estrangeira. Seu foco, portanto, são os textos específicos da academia, tais como resenha, resumo, ensaio, artigo científico, pôster, comunicação em evento, oficina, palestra, mesa redonda, etc.”, conjuntamente à disciplina antes mencionada, a carga horária da disciplina da área de tradução contém sessenta horas teóricas, sendo nomeada “Língua Espanhola: Estudos em Tradução”, descrita, na ementa, da seguinte forma: “Essa disciplina proporciona debates sobre a identificação e caracterização dos princípios e procedimentos usados pelo tradutor no ato de tradução; os modelos de tradução, análise crítica de traduções consagradas a partir de comparações com o texto original; classificação dos processos e procedimentos usados pelo tradutor.”; além das disciplinas já mencionadas, existe uma focada nas metodologias próprias dos professores de espanhol, chamada “Metodologia de Ensino da Língua Espanhola”, com caráter teórico-prático, possuindo 45 horas teóricas e 15 horas práticas na carga horária, definida na ementa da seguinte forma: “Esta disciplina proporciona o debate sobre a construção do perfil do professor de língua estrangeira enquanto cidadão ético, crítico, político e reflexivo; a formalização das metodologias, abordagens, métodos e técnicas de ensino de língua estrangeira, tanto no que tange ao ensino das quatro habilidades quanto ao ensino para fins específicos; análise, escolha e produção de material didático; planejamento de curso e de aulas; avaliação.” e; finalmente, uma disciplina que possui trinta horas práticas na carga horária, “Pipe 5: A Criatividade no Ensino de Língua Espanhola – Técnicas de Ensino e Interações”, com a seguinte ementa: “Levantamento de técnicas de ensino e interação em sala de aula”.

No sexto período, o currículo contem exclusivamente disciplinas da especialidade de língua espanhola, três destas fazem parte do “Componente Curricular do Módulo de Língua Espanhola”, possuindo sessenta horas teóricas na carga horária cada. Assim, o aluno do sexto período tem uma variada oferta de componentes para fazer sua escolha. Dentre as opções de Componentes Curriculares do Módulo de Língua Espanhola encontram-se as seguintes disciplinas: “Língua Espanhola: Civilização e Cultura Espanhola”, “Civilização e Cultura dos Povos de Língua Espanhola”, “Língua Espanhola: Leitura Instrumental”, “Língua Espanhola:

Leitura Para Fins Acadêmicos”, “Ensino de Língua Espanhola e as Novas Tecnologias”, “Língua Espanhola em Contexto Empresarial”, “Língua Espanhola: Fonética e Fonologia”, “Morfossintaxe Da Língua Espanhola”, “Língua Espanhola: Avaliação da Aprendizagem”, “Língua Espanhola: Estudos Descritivos”, “Língua Espanhola: Análise e Elaboração de Material Didático”, “O Contexto Global do Ensino de Língua Espanhola”, “Língua Espanhola: Introdução aos Estudos sobre Identidade”, “Língua Espanhola: Tradução de Artigos Acadêmicos”, “Tradução de Quadrinhos: Espanhol/Português”, “Tradução de Filmes – Espanhol/Português”. Também o aluno do sexto período tem a possibilidade de escolher uma disciplina dos “Componentes Curriculares do Módulo de Literatura Espanhola” que, do mesmo modo, consta de sessenta horas teóricas na carga horária, sendo as seguintes disciplinas como opções: “Literatura Espanhola: Siglo de Oro”, “Literatura Espanhola: Da Idade Média ao Renascimento”, “Literatura Espanhola Contemporânea”, “Literatura Espanhola: Teatro Espanhol”, “Literatura Espanhola: Literatura e Cinema Espanhol”, “Literatura Espanhola e Cinema Espanhol: Luis Buñuel e o Surrealismo”, “Literatura Espanhola: Teatro Espanhol e Sua Representação Textual”, “Literatura Espanhola: Teoria do Ensaio”, “Literatura Hispano-americana: da Colonização ao Século XVIII”, “Literatura Hispano-americana: do Romantismo as Vanguardas”, “Literatura Hispano-americana Contemporânea”, “Literatura Hispano-americana: a ‘Arte de Convergência’ de Octavio Paz”, “Literatura Hispano-americana: Brasil e Argentina – Convergências Culturais”, “O Conto Hispano-American”, “Literatura Hispano-americana: Vanguardas Históricas do Século XX”, “Literatura Hispano-americana: A Novela”, “Literatura Hispano-americana: A Poesia Mexicana Moderna – ‘Confluências’”, “Literatura e Expressão Espanhola: Estudos Interartes – Origens e Casos”, “Literatura Espanhola e Hispano-americana no Brasil: Estudos Comparados”, as outras duas disciplinas que encerram o currículo do sexto período estão voltadas ao ensino, sendo estas: “Metodologia de Ensino do Espanhol para Fins Específicos”, que se perfila como uma disciplina de caráter teórico-prático, contando com 45 horas teóricas e 15 horas práticas na carga horaria, definida na ementa da seguinte forma: “Estudo sobre o ensino de Língua Espanhola para fins específicos. Os componentes que constituem esta disciplina abrangem o conceito de ensino para fins específicos e as concepções de língua e aprendizagem subjacentes a essa abordagem.” e; “Pipe 6: Investigando Necessidades e Interesses Para o Ensino de Língua Espanhola em Diferentes Contextos” que diferente da disciplina anterior, possui caráter totalmente prático, contando com trinta horas práticas na carga horaria, definindo-se na ementa da seguinte maneira: “Desenvolvimento de práticas educativas sobre a elaboração e aplicação de questionários e

entrevistas em contextos pré-determinados para o levantamento de ‘necessidades e interesses’ e organização, estruturação e compilação dos dados para a redação de relatório.”. Totalizam-se, no final do terceiro ano de estudos, 570 horas de estudo da especialidade de língua espanhola, sendo estas 480 horas teóricas e 90 horas práticas.

No sétimo período, o currículo está composto por quatro disciplinas específicas da língua espanhola, correspondendo, duas delas, ao “Componente Curricular do Módulo de Literatura Espanhola”. Estas disciplinas, como já antes mencionadas, possuem sessenta horas teóricas na carga horária, além da disciplina “Estágio Supervisionado em Língua Espanhola”, disciplina teórico-prática, com 30 horas teóricas e 45 horas práticas, definida na ementa como: “Abordagem comunicativa para o ensino da habilidade de expressão oral em língua espanhola, planejamento, avaliação e regência de sala de aula.” e; uma disciplina de caráter prático, “Pipe 7: Língua Espanhola - Seminários de Práticas Educativas”, a qual está composta por 45 horas práticas na carga horária, com a seguinte ementa: “Finalização do Projeto Integrado de Práticas Educativas, com apresentação dos resultados das atividades realizadas nos diferentes momentos de seu desenvolvimento, proporcionando a articulação teórico-prático-pedagógica, pretendida entre os conhecimentos específicos da área, os conhecimentos educacionais e pedagógicos e o fazer profissional desses professores.”.

No oitavo, último período do curso, os alunos realizam quatro disciplinas específicas do espanhol, sendo uma destas do “Componente Curricular do Módulo de Literatura Espanhola” com sessenta horas teóricas na carga horária. Podemos encontrar também a disciplina teórico-prática “Estágio Supervisionado de Práticas em Tradução”, possuindo 30 horas teóricas e 60 horas práticas na carga horária, definida na ementa da seguinte maneira: “O reconhecimento da possibilidade de se realizar um processo tradutório levando-se em conta a literalidade e os efeitos de sentido. O papel do tradutor em um mundo globalizado. Os diversos tipos de tradução e as novas tecnologias. Os tipos de pesquisa empírica e o trabalho metodológico.”. A disciplina de estágio em língua espanhola também acontece no oitavo período, intitulando-se “Estágio Supervisionado em Língua Espanhola 2”, o qual possui 30 horas teóricas e 60 horas práticas, definindo-se na ementa da seguinte maneira: “Abordagem comunicativa para o ensino da habilidade de leitura e compreensão de textos escritos em língua espanhola, planejamento, avaliação e regência de sala de aula.” e; finalmente, a disciplina “Estágio Supervisionado em Espanhol Para Fins Específicos”, o qual consta de 15 horas teóricas e de 60 horas práticas na carga horária, definida pela ementa como: “Formação de

professores de língua espanhola para atuação em ensino de Espanhol para fins específicos.”¹⁶. Desta forma, vemos como totalizam-se 735 horas da especialidade da língua espanhola durante o último ano de Curso de Letras Habilitação em Espanhol, correspondendo 465 à horas teóricas e 270 à horas práticas.

Tabela 2 – Horas de estudo de disciplina especializadas em espanhol durante o curso.

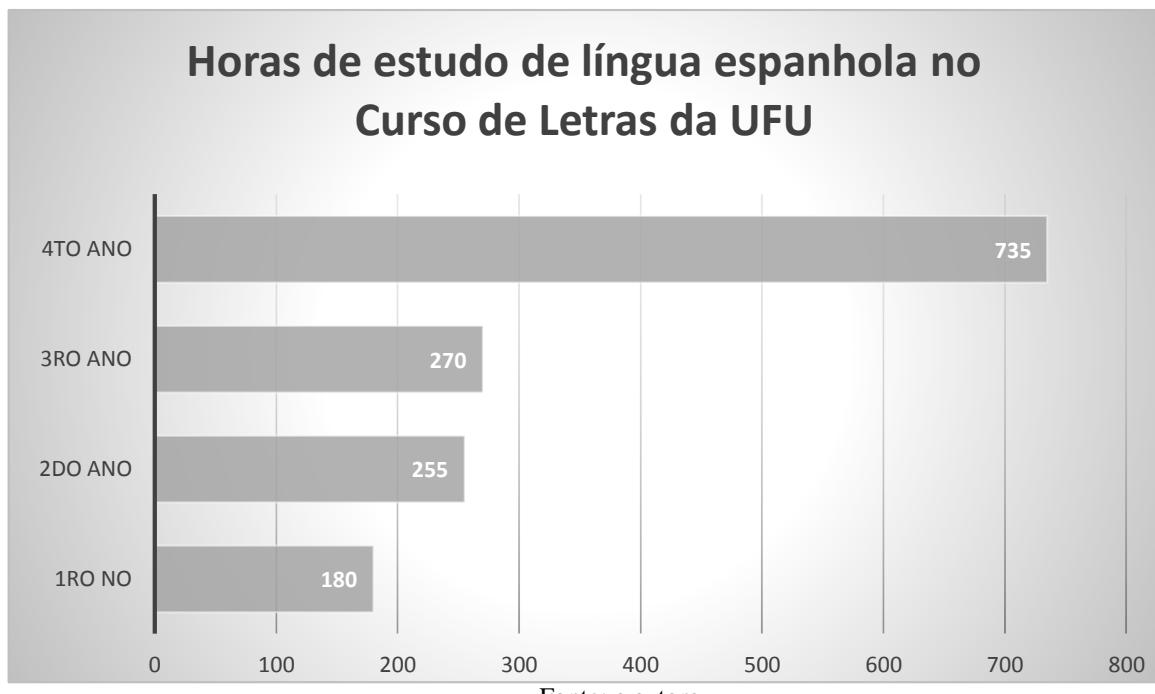
	1º	2º	3º	4º	5º	6º	7º	8º
Horas teóricas	60	60	60	180	225	225	150	315
Horas práticas	30	30	15	-	45	45	90	180
Total	90	90	75	180	300	270	240	495

Fonte: a autora.

Na Tabela 2, podemos observar o resumo da quantidade de horas de disciplinas específicas de espanhol oferecidas durante o Curso de Letras. Vemos que a tendência é aumentar o número de horas de estudo da língua estrangeira, isto evidencia-se ao apreciarmos que no primeiro período do curso o total estudado é de 90 horas, enquanto no oitavo e último período aumenta para 495 horas, ou seja, quintuplicando o número de horas.

¹⁶ Extraído do site http://www.ileel.ufu.br/letras/?page_id=12830 em 20 de outubro de 2015.

Figura 2 – Horas de estudo de língua espanhola por ano cursado no Curso de Letras da UFU.



Fonte: a autora.

Após a confirmação da aprovação do projeto de pesquisa por parte do CEP, os alunos selecionados, que preencheram o questionário e atenderam aos requisitos da pesquisa, foram contatados para, desse modo, definir um horário conveniente para os mesmos, para a realização e aplicação das atividades no LABLING.

Nesse sentido, antes de começar as atividades orais e escritas propostas em cada um dos períodos, entregamos uma folha com o TCLE, que explica detalhadamente os propósitos e objetivos da pesquisa. Assim, os estudantes que demonstraram concordância com a proposta e que se interessaram em fazer parte da pesquisa, ficaram envolvidos com a mesma, logo após a assinatura do TCLE, com a qual confirmaram seu conhecimento, a respeito da pesquisa, e sua disposição em participar.

Em todos os períodos, primeiramente, foi realizada a aplicação das atividades orais, utilizando, para isso, o espaço oferecido pela UFU já mencionado anteriormente, o LABLING. Os enunciados das atividades foram projetados em uma tela e explicados até que todas as dúvidas dos participantes fossem respondidas, antes de começar a aplicação.

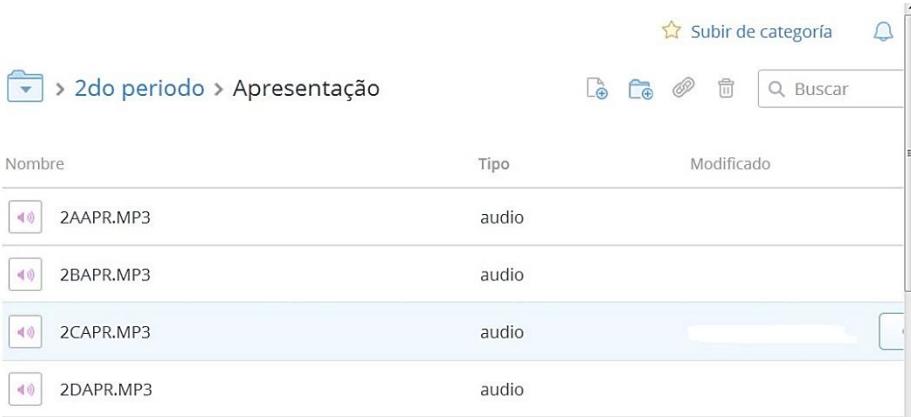
A ordem de aplicação das atividades orais foi a seguinte: primeiro a apresentação, logo a leitura, diálogo de duas pessoas e finalmente um debate no qual participaram os quatro participantes do período juntos. Uma vez finalizada a parte das atividades orais, fez-se a entrega do documento com as atividades escritas para cada um dos participantes. Essa etapa consistiu

em duas perguntas de opinião, uma descrição e, finalmente, uma carta que os estudantes deviam escrever a partir de uma situação hipotética proposta (Ver apêndice B).

Imediatamente depois das seções de coleta dos dados, todos os arquivos de áudio foram arquivados na plataforma virtual *Dropbox*, uma plataforma virtual gratuita, que brinda o conforto de poder acessar os arquivos de qualquer computador, desde que se tenha uma conexão com a Internet.

A nomenclatura para os arquivos obedeceu a seguinte lógica: O primeiro número indica qual o período onde foi aplicada a atividade (2, 4, 6 ou 8), seguido por uma letra (A, B, C ou D) que indica qual o participante e, finalmente, as três primeiras letras que indicam qual a atividade que se realizou: Apresentação, Leitura, Diálogo ou Debate (APR, LEI, DIA ou DEB). Na Figura 3, a seguir, pode-se apreciar um exemplo de como foram nomeados os arquivos de áudio da atividade *Apresentação* dos participantes do segundo período.

Figura 3 – Armazenamento na plataforma Dropbox.



Nombre	Tipo	Modificado
2AAPR.MP3	audio	
2BAPR.MP3	audio	
2CAPR.MP3	audio	
2DAPR.MP3	audio	

Fonte: a autora.

O próximo passo foi o armazenamento dos documentos das atividades de produção escrita, mas para isso foi preciso digitalizar os documentos que estavam em formato impresso. Para esse propósito, foi utilizado o *Scanner Planetário* da Biblioteca da UFU. O referido equipamento possibilita o envio dos documentos digitalizados até o correio eletrônico ou que sejam salvos, por exemplo, num *pen-drive*.

Figura 4 – Scanner Planetário. Biblioteca da UFU, Campus Santa Mônica.



Fonte: a autora.

A próxima figura, Figura 5, ilustra o resultado da digitalização de uma das produções escritas dos estudantes, após o procedimento de escaneamento ou digitalização.

Figura 5 – Amostra de digitalização das atividades de produção escrita.

LEI DO DIREITO AUTORAL
Todos os direitos reservados e protegidos
pela Lei 9.610/1996.

ESTADO MINEIRO/UFU
Todos os direitos reservados e protegidos
pela Lei 9.610/1996.

Este arquivo não pode ser reproduzido
sem a devida autorização das autoridades
acadêmicas que o criaram ou de suas
equivalentes.

PRODUÇÃO ESCRITA

I. Lee atentamente el texto que se presenta a continuación y luego responde a las preguntas que se formularán. Recuerda que es importante tu fundamentación, utiliza todo el espacio destinado para las respuestas:



La lengua no es un negocio, pero a menudo se la trata como tal, y entre algunas corporaciones españolas, por ejemplo, cuando la metáfora de compararla con el petróleo. Los negocios vinculados a la lengua – traducciones, editoriales, comunicaciones, audiovisuales – constituyen alrededor del 15 % del PBI español. El Instituto Cervantes, como difusor de la cultura hispanohablante y la enseñanza del idioma, es un dispositivo central de esa configuración mercantil. Y la atribución de la capacidad normativa a la Real Academia española convierte a la variedad considerada estándar por la misma como norma para todas las industrias vinculadas a la lengua. Por ejemplo, cuando cualquier latinoamericano usa el procesador de textos Word y elige su variedad –español de Argentina–, el diccionario contra el cual confronta la corrección está producido por la RAE y no reconoce los usos habituales de la variedad argentina. Eso implica dos cuestiones: insigualdad lingüística del escritor local y cuantiosos convenios económicos entre la RAE y Microsoft", ejemplifica María Pía López, directora del Museo de la Lengua y una de las promotoras del documento "Por una soberanía idiomática", que firmaron medio centenar de lingüistas, escritores y académicos argentinos el 17 de septiembre en Página12. En él, proponen, entre otras medidas, la creación de un Instituto Borges (¿un Cervantes argentino?) –que sirva para plantear el discurso político de la lengua– y para más adelante la creación de una Asociación Latinoamericana de Lengua.

Extruido de: http://www.revistaenlacar.com/ideas/VI-Congreso-de-la-Lengua-Fernando-Vallejo-Sergio-Ramirez_0_101399668.html

Preguntas:

1. ¿Qué te parece que un organismo europeo fije las normas del español modelo? ¿Por qué? (fundamenta tu respuesta)

(Bartolomé) Muy conveniente para con sus padres e ideologías. Yo creo que es peligroso hasta cierto punto crear padres para las cosas en general. Por eso soy a favor de la diversidad, aún pienso que la diversidad genera igualdad a la medida que respeta los puntos de vista involucrados. Entonces pienso que padres son malos, incluyendo y principalmente padres de los que forman la historia humana de acuerdo con su punto de vista "totalitario"

2. ¿Qué tipo de español crees que debería ser enseñado en los países no hispanohablantes americanos como Brasil y Estados Unidos? ¿Tienes preferencia por alguna variante en especial? ¿Por qué? (fundamenta tu respuesta)

Un español que respechase las variedades locales con sus idiosincrasias respectivas. Tengo preferencia (como) por las variantes locales que son habladas por pueblos indígenas, porque me gusta car el punto de vista de los pueblos necesitados, pero que sin embargo tiene una cultura ancestral y de respeto para con la Madre Naturaleza.

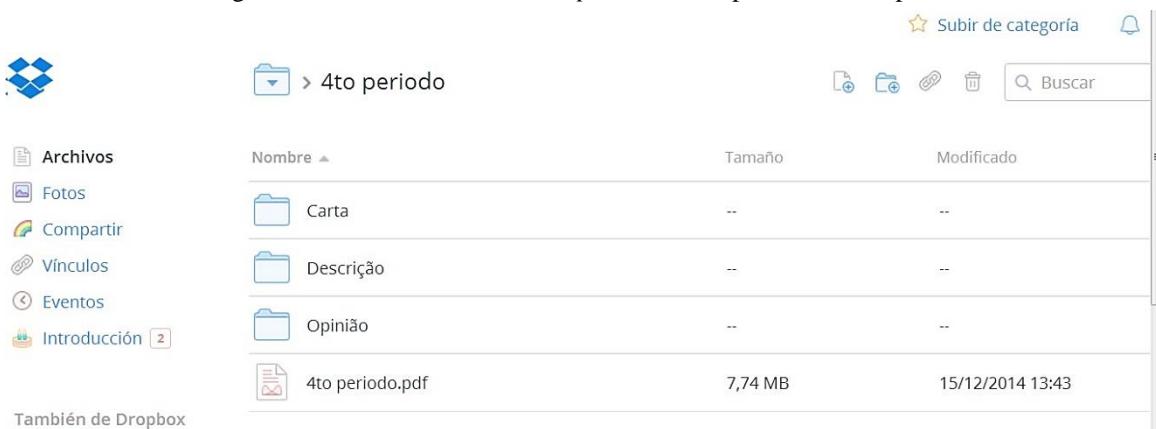
II. Haz una descripción de la UFU, también puedes basarte en el Curso de Letras/Español, puedes tomar en cuenta todos los aspectos que estimes pertinentes (estructura, espacios, cursos ofrecidos, características de los alumnos, disciplinas, profesores, sistema de ingreso, entre otros)

Primero que la UFU tiene intención de producir saberes y conocimiento. Esto es bueno y esencial para mejorar la sociedad. Pero por otro lado falta mejorar con relación a adquirir energía solar y a partir de esta mejoría, también avanzar de desmantelar sus pocas áreas verdes y construir edificios verticales movidos por energía solar. Esto es solamente un ejemplo de los problemas que veo. Cuanto a el Curso de Letras en general, me gusta, principalmente literatura y posibilidades de interdisciplinarias. Un problema que veo es el hecho de que el curso de Letras, solamente produce a tiene la motivación de producir profesores y pesquisadores. Yo diría escritores de

Fonte: a autora.

A digitalização dos documentos com as atividades de produção escrita foi feita somente separando os arquivos pelo período correspondente (segundo, quarto, sexto e oitavo período). Imediatamente, esses arquivos foram submetidos à plataforma do *Dropbox*, assim como foi feito com os arquivos de áudio, armazenando os arquivos em pastas independentes.

Figura 6 – Armazenamento do *corpus* escrito na plataforma Dropbox.



Fonte: a autora.

Finalmente, já com os arquivos disponibilizados na plataforma virtual, começou-se o processo de transcrição dos arquivos sonoros, tomando em conta, para essa tarefa, a transcrição literal do áudio, com o intuito de conferir se as palavras estavam sendo pronunciadas da maneira correta ou se, pelo contrário, elas estariam sendo pronunciadas com algum tipo de anomalia que, mais adiante, iriam configurar a idiossincrasia do “erro” à qual pertence.

Para esse propósito de transcrição, a lógica para denominação dos arquivos transcritos foi a mesma descrita anteriormente (número por período, letra por participante e abreviação das três primeiras letras da atividade). No momento de salvar os arquivos, empregamos dois formatos diferentes *DOC*, para posterior anotação e etiquetagem do *corpus*, e em *TXT*, para poder processar o *corpus* de estudo com o software *Word Smith Tools*.

A Figura 7 ilustra uma vista parcial da transcrição de um Debate, salvo em formato *TXT*:

Figura 7 – Amostra da transcrição da atividade oral de Debate.

2ADEB - Bloco de notas

Arquivo Editar Formatar Exibir Ajuda

A: Yo estoy de acuerdo con Mujica porque yo también pienso que el problema no es la marijuana y sí el narcotráfico que incluso está agregado con muchas veces con la policía / con instituciones políticas / entonces una ley que regulamente el consumo es una buena solución creo que como él dijo no es la solución inmediata pero creo que es el primer paso

B: Yo también concordo con Mujica que no es una cuestión de principios son los pequeños traficantes los pobres que mueren con el narcotráfico y esta cuestión de salud pública que tiene que ser sancionada: Yo creo que concordo también pero ellos comparan con el tabaco y el alcohol pero no sé si será el misma cosa porque es una bagunza isso / menores de dieciocho años consiguen fácilmente y será que bien regulamentada misma esa ley

D: Sí, yo también concordo com Muyica mas pero yo creo que há muitas coisas a serem feitas antes de que se torne legalizada como la melhor estrutura para las personas porque es una questao de personas y no de un todo mas creo que esta ley tiene que ser muy bien planejada porque sino como bien que disse será uma bagunza

C: Y será que acaba con el narcotráfico mesmo mesmoD: Entao, eu acho que no, que continuaráA: Sí porque tene cocaína, tene craqui y otras drogas

C: Pero continuará vendendoD: Pero continuará mesma coisa

C: Continuará mesmo sin la marijuanaD: Será más fácil de conseguir

A: Sí pero la marijuana es como si fosse el motor que conduce al tráfico/ se compra mais marijuana do que cocaína do que craqui / heroína / entonces yo creo que será un gran golpe para el narcotráfico

C: Yo creo que el gobierno tiene muy interesse en eso por questiones de impuestos

A:Sí / y esos impuestos si bien conducidos podrán mejorar las condiciones de educación / de salud / de trabajo / podrán ser bien investidos esos recursosD: En uruguá sí mas en Brasil yo hago que no / porque yo penso que yo no temos un governo tão cabeça aberta como en Uruguay / yo penso así y acho que A: También Brasil es un país muy maior, muy grande en territorioD: Muito difícil controlar/ há muitos problemas / nuestras vidas con muchas cosas no só com las drogas / entao acho que será mais difícil en Brasil B: Sí / entonces acreditan que es mejor liberar todos los tipos de drogas

A: No / no ahora / por lo menos no ahora

C: Acho que la población ainda nao tem uma maturidá para enfrentar isso

Fonte: a autora.

Após o exame de qualificação, dentre algumas anotações feitas pela banca avaliadora e para evitar qualquer tipo de confusão na leitura dos exemplos que mais adiante aparecerão, decidimos acrescentar, à transcrição ortográfica, a acentuação das palavras que não fossem monossilábicas, para, desta maneira, poder diferenciar a acentuação no texto escrito e no programa que irá processar os dados e, desta forma, saber quando exista uma palavra pronunciada de forma errada. A seguir, a Figura 8 traz uma amostra do enriquecimento da transcrição já feita:

Figura 8 – Amostra do enriquecimento da transcrição ortográfica na atividade de Debate no sexto período.

C: Bueno / yo cuándo piéndo sobre éso / no sé que pensár tambiéñ / no sé si la legalización sería buéna pára Brasil o no / yo créo que la curiosidá <MOR[curiosidad]> de los adolescentes con la legalización sería mayór / no sé / ótras véces yo piéndo que no pórque éso tambiéñ viene de cáda persona / cómo las personas que fuma <MOR [fuman]> y ótras que no yo créo que tal vez no sería Una influyén <LEX [influencia]> úna un inflúyo <LEX [influencia]> pára que la mayoría de la población experimentaría la marijuána <FON : ING [marihuana]>

A: Yo péndo <MOR [pienso]> / péndo <MOR [pienso]> que la limitación hác con que <SIN : OM [hace que]> los órganos de salúd pública sí / puédan estar cérc de las personas que ultrapásen <LEX [sobrepasen]> ese límite <FON [límite]> sí / controlár eso / lo que no es posíble cómo ocórr <MOR [ocurre]> en Brasiu <FON [Brasil]> / cargaméntos de cién toneládas de drógas tódos los días y várías personas que se muéren por hora

C: Péro la marijuána <FON : ING [marihuana]> / yo créo que no es tan perjudicíal / tiéne ótras drógas mucho más fuértes que siémpre van a estár ahí

B: También / péro el alcojól <FON [alcohol]> y el cigarrillo es legál / péro aún hay mucho contrabando de éso / o séa / es legál y no acabó los problémás

A: No / no / creíto lo lavo pérro inhibir el consumo de alcohol / no es ciumento

PAGE 1 OF 2 703 WORDS SPANISH (CHILE) 108 %

Fonte: a autora.

A próxima figura, Figura 9, traz a organização do segmento *Apresentação*, do *corpus* oral do segundo período, no sentido de ilustrar o armazenamento em pastas dos arquivos de áudio (MP3), as transcrições (DOC) e correspondentes versões em TXT para leitura com as ferramentas do WST. Seguindo os mesmos critérios, todos os arquivos que compõem o *corpus* de estudo foram armazenados em pastas e subpastas e compartilhados no *Dropbox* entre orientador e orientanda.

Figura 9 – Armazenamento da atividade oral Apresentação, segundo período, no Dropbox.

Nombre	Tipo	Modificado
2AAPR.MP3	audio	12/12/2014 10:41
2BAPR.MP3	audio	12/12/2014 10:41
2CAPR.MP3	audio	12/12/2014 10:41
2DAPR.MP3	audio	12/12/2014 10:41
2AAPR.doc	documento	9/3/2015 08:34
2BAPR.doc	documento	9/3/2015 08:34
2CAPR.doc	documento	9/3/2015 08:35
2DAPR.doc	documento	9/3/2015 08:34
2AAPR.txt	documento	9/3/2015 08:35
2BAPR.txt	documento	9/3/2015 08:35
2CAPR.txt	documento	9/3/2015 08:35
2DAPR.txt	documento	9/3/2015 08:35

Fonte: a autora.

Quanto ao *corpus* escrito, este foi primeiramente digitalizado, como dito anteriormente, para depois ser digitado e, portanto, convertido ao formato de texto digital em *DOC* e posterior conversão a *TXT*. Nessa oportunidade, utilizamos os mesmos critérios para nomeação dos arquivos: número, segundo o período em curso (2, 4, 6 ou 8); letra, segundo o informante (A, B, C ou D) e, finalmente, as três primeiras letras da atividade realizada (OPI, DES ou CAR).

Finalizada a etapa de transcrição de arquivos de áudio, digitação de arquivos digitalizados e conversão do formato *DOC* para *TXT*, os arquivos de áudio originais foram apagados e o material em papel das atividades de produção escrita foi devolvido a cada um dos participantes da pesquisa, conforme descrito no TCLE e no projeto aprovado no CEP.

Logo após o descarte dos arquivos originais, e uma vez já compilado e organizado nosso *corpus* de estudo oral e escrito, restaria saber suas dimensões. Para esse propósito, utilizamos a ferramenta *WordList* do programa WST, na versão 6.0 (SCOTT, 2012). A partir dos dados gerais de ambos os *corpora*, oral e escrito, obtivemos um total de oitenta arquivos em formato *TXT* que, processados pela ferramenta *WordList*, lançou um total de 20.385 palavras totais ou Itens (*tokens*) e 2.961 palavras diferentes ou Formas (*types*). Com a extensão do *corpus* de estudo definida, tivemos a base para a etapa seguinte: o trabalho de identificação

da idiossincrasia dos “erros” e a sua etiquetagem, para posterior quantificação, descrição contrastiva e análise qualitativa das ocorrências. A Tabela 3 apresenta a extensão do *corpus* de estudo:

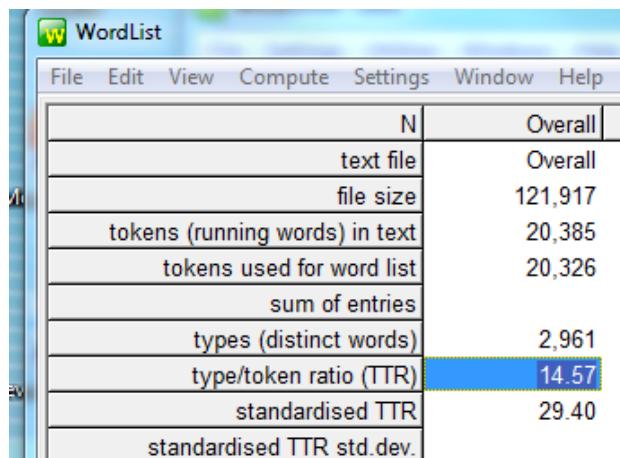
Tabela 3 – Extensão do *corpus* de estudo.

	Documentos	Itens (Tokens)	Formas (Types)	Razão Forma/Item (%)
Corpus Oral	37	15.132	2.086	13,79
Corpus Escrito	43	5.253	1.379	22,55
TOTAL	80	20.385	2.961	14,57

Fonte: a autora.

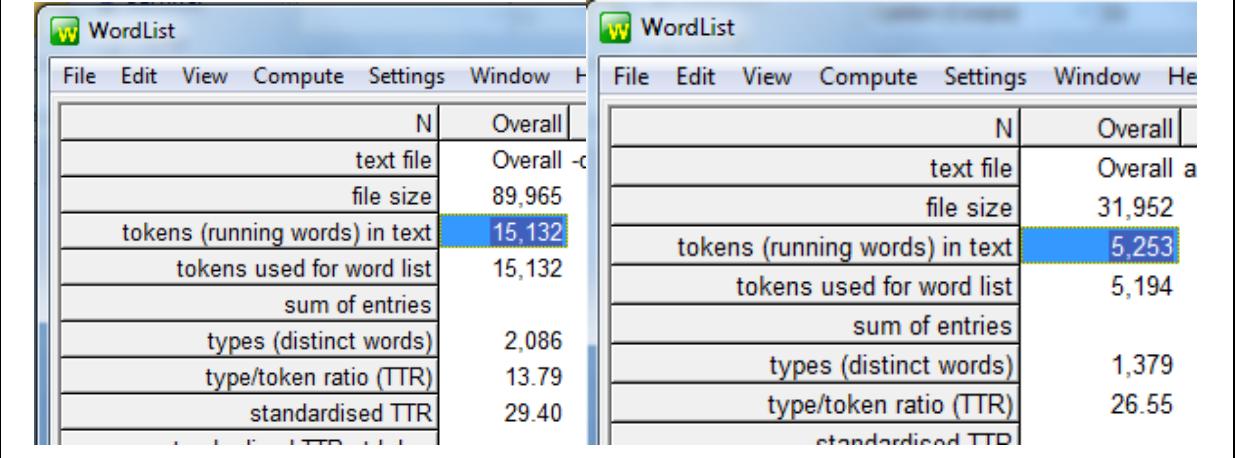
A próxima figura, Figura 10, apresenta uma vista parcial dos dados obtidos com a ferramenta *WordList* do WST:

Figura 10 – Dimensão do *corpus* no *WordList*.



Fonte: a autora.

A primeira estatística que tivemos interesse em fazer foi a que nos indicaria o número total das palavras que o *corpus* oral contém. Para esse propósito, a ferramenta do programa utilizada foi *WordList*, usada para listar o número total de palavras. A Figura 11 apresenta os dados estatísticos do *corpus* oral e do *corpus* escrito:

Figura 11 – Dados estatísticos do *corpus* oral e escrito no WST.


The image shows two identical instances of the WordList software interface side-by-side. Both windows have a blue header bar with the 'WordList' logo and a menu bar: File, Edit, View, Compute, Settings, Window, and Help. The main area is a table with two columns: 'N' and 'Overall'. The data in the table is as follows:

	N	Overall
text file	Overall	89,965
tokens (running words) in text	15,132	31,952
tokens used for word list	15,132	5,253
sum of entries	2,086	5,194
types (distinct words)	13.79	1,379
type/token ratio (TTR)	29.40	26.55
standardised TTR		

Fonte: a autora.

Quanto aos códigos utilizados na transcrição, nos baseamos nos códigos de transcrição propostos pelo grupo de pesquisa CORELE para a realização do seu *corpus* linguístico de 2014, adaptando-os em razão dos objetivos da presente pesquisa. A seguir a lista de códigos utilizados para a transcrição e etiquetagem de “erros”:

Tabela 4 – Códigos de transcrição.

Códigos de transcrição	
<> - TAGS	Para indicar os níveis em que ocorreram as inadequações que irão ser estudadas: fonético, ortográfico, morfológico, sintático e lexical.
(())	Palavra ou fragmento incompreensível.
[]	A palavra correta, depois da inadequação transcrita
/	Pausa curta
//	Pausa mais prolongada
-	Reinícios, auto interrupções
“h”	Na posição final da sílaba - aspiração
z	Pronuncia sibilante da /s/ e da /z/. Por exemplo, pronunciar ‘coza’ assim [cosa].

Tabela 5 – Níveis de análise das inadequações.

Níveis de análise das inadequações	
<FON>	Fonética
<MOR>	Morfológica
<LEX>	Lexical
<SIN>	Sintática
<MOR:SIN>	Morfológico e Sintático
<LEX:SIN>	Lexical e Sintático
:OM	[Colocar aqui a palavra que sobra]
:AD	[Colocar aqui a palavra que falta].
:ING>	Quando tem influência do inglês.
:FRN>	Quando tem influência do francês.

Em primeira instância, realizou-se a etiquetagem de um recorte do *corpus*, especificamente do segundo período, o qual seria utilizado para apresentar resultados preliminares no exame de qualificação. Após a qualificação, continuou-se o processo de etiquetagem de ambos os *corpus*, utilizando, para esse propósito, os códigos mencionados acima.

Os tipos de erros a serem etiquetados seriam: fonológicos (no *corpus* oral), ortográficos (no *corpus* escrito), morfológicos, sintáticos e lexicais. Para fazer a marcação do erro, após a sua identificação, colocou-se entre <> o código correspondente para o tipo de erro (FON, ORT, MOR, SIN ou LEX), em seguida, entre [] colocou-se a forma correta ou esperada para aquele erro. Foi possível, também, identificar mais de um erro simultâneo. Nesse caso, o procedimento foi colocar entre os dois tipos de erros, o signo de dois pontos (ex. MOR:SIN).

Para os erros de tipo sintático, utilizou-se os códigos OM, para indicar quando

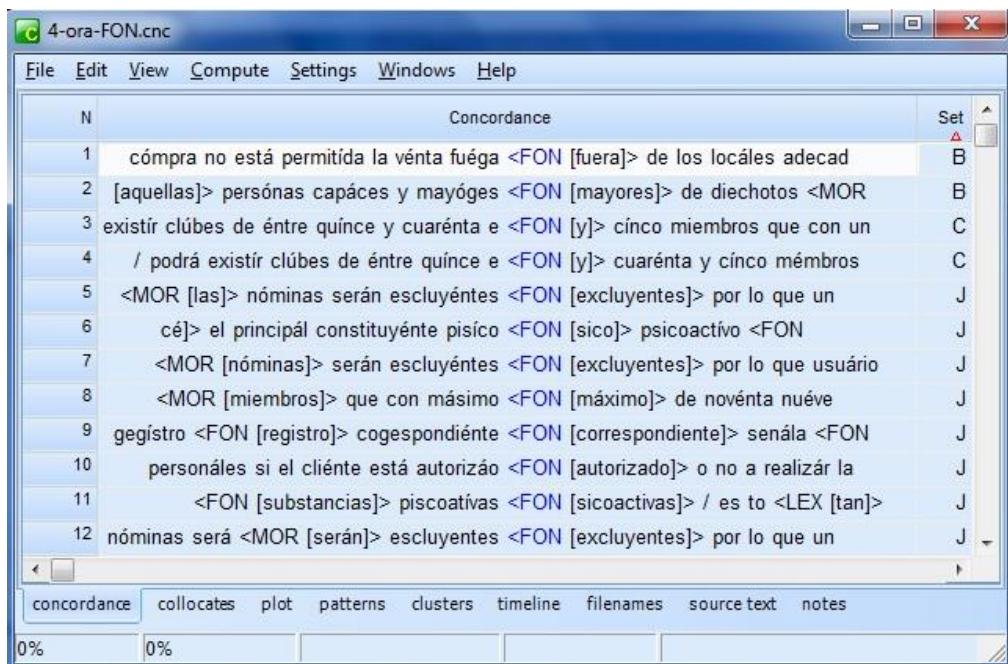
existia uma palavra que devia ser apagada e, também, o código AD, o qual indicara que era preciso adicionar mais uma palavra para que a frase fosse sintaticamente correta.

Além dos códigos já mencionados, incluímos os códigos ING, para quando se identificasse um erro que possivelmente tivesse sua origem ou influência na língua inglesa, e o código FRN, para erros que pudessem ter influência da língua francesa. Isto é pertinente se considerarmos que a formação nos períodos iniciais do curso tem disciplinas das línguas antes mencionadas para além do espanhol.

O processo de etiquetagem do *corpus* passou por várias instâncias de correção durante o segundo semestre de 2015, foi preciso rever junto com o professor orientador, até chegar num ponto de aprimoramento que permitisse o processamento dos dados com as ferramentas de análise de *corpus* linguístico.

Uma vez etiquetado o *corpus* linguístico, adicionamos mais uma categoria de análise. Identificamos a categoria gramatical de cada um dos “erros” identificados no *corpus*. Para esse propósito, utilizamos a ferramenta *Concord*, preenchendo a coluna “set” que se encontrava disponível para adicionar mais algum dado.

Figura 12 – Amostra da identificação da categoria gramatical dos erros no quarto período do *corpus* oral.



The screenshot shows a window titled '4-ora-FON.cnc' with a menu bar: File, Edit, View, Compute, Settings, Windows, Help. The main window is a 'Concordance' table with columns: N (number), Concordance (text), and Set (category). The Set column contains codes: B, B, C, C, J, J, J, J, J, J, J, J. The text in the Concordance column is in Spanish and discusses various grammatical errors and their contexts.

N	Concordance	Set
1	cómpra no está permitida la vénta fuéga <FON [fuera]> de los locáles adecad	B
2	[aquellas]> persoñas capáces y mayóges <FON [mayores]> de diechotos <MOR	B
3	existir clúbes de éntre quínce y cuarénta e <FON [y]> cinco miembros que con un	C
4	/ podrá existir clúbes de éntre quínce e <FON [y]> cuarénta y cinco mémbros	C
5	<MOR [las]> nóminas serán escluyéntes <FON [excluyentes]> por lo que un	J
6	cé]> el principál constituyénte pisico <FON [sico]> psicoactivo <FON	J
7	<MOR [nóminas]> serán escluyéntes <FON [excluyentes]> por lo que usuário	J
8	<MOR [miembros]> que con másimo <FON [máximo]> de novénta nuéve	J
9	gegistro <FON [registro]> cogenpondiénte <FON [correspondiente]> senála <FON	J
10	personáles si el clíente está autorizáo <FON [autorizado]> o no a realizar la	J
11	<FON [substancias]> piscoatívas <FON [sicoactivas]> / es to <LEX [tan]>	J
12	nóminas será <MOR [serán]> escluyentes <FON [excluyentes]> por lo que un	J

Below the table, there is a navigation bar with tabs: concordance, collocates, plot, patterns, clusters, timeline, filenames, source text, notes. The 'concordance' tab is selected. At the bottom, there are progress bars for '0%' and '0%'.

Fonte: a autora.

Quanto aos códigos utilizados para a identificação das categorias gramaticais, nos baseamos nos códigos propostos no trabalho da aluna de iniciação científica, Silva (2015), adaptado segundo os objetivos da presente pesquisa.

Tabela 6 – Códigos para a identificação das categorias gramaticais dos “erros” do *corpus* linguístico.

PARTES DA ORAÇÃO	
Artigo	(A)
Adjetivo	(J)
Advérbio	(B)
Conjunção	(C)
Interjeição	(I)
Preposição	(P)
Pronome	(R)
Verbo	(V)
Substantivo	(S)
Contração	(N)

Fonte: a autora.

Para continuar com o processo de análise de dados, primeiramente utilizamos a ferramenta de *WordList* do programa *Word Smith Tools 6.0*, o qual nos permitiu ter uma visão geral de como ficava configurado o *corpus*, quanto à sua dimensão. Para este propósito definimos qual era a dimensão total do *corpus* oral e do *corpus* escrito, organizando-os por *subcorpora* por período e por atividade. A seguir, Figura 13, uma amostra dos dados do quarto período do *corpus* oral.

Figura 13 – Amostra da dimensão do *subcorpora* do quarto período do *corpus* oral.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/oke standardis ratio
1	Overall	67,260	4,451	4,451	1,059	23.79
2	4ALEI - ETQ.txt	9,722	571	571	296	51.84
3	4BLEI - ETQ.txt	9,010	566	566	286	50.53
4	4CLEI-ETQ.txt	7,964	568	568	281	49.47
5	4DLEI -_ETQ.txt	8,902	575	575	286	49.74
6	4ADIA-ETQ.txt	6,080	476	476	191	40.13
7	4BDIA - ETQ.txt	5,356	383	383	188	49.09
8	4ADEB - ETQ.txt	12,762	809	809	348	43.02
9	4AAPR-ETQ.txt	1,256	71	71	50	70.42
10	4BAPR-_ETQ.txt	1,596	115	115	72	62.61
11	4CAPR-_ETQ.txt	2,860	222	222	125	56.31
12	4DAPR-_ETQ.txt	1,752	95	95	68	71.58

Fonte: a autora.

Na Figura 13, podemos ver uma amostra dos dados correspondentes à quantidade de arquivos utilizados, o número de *tokens*, *itens* e da razão forma/item.

Outro nível de análise dessa pesquisa, foi o das palavras mais recorrentes de ambos *corpora*, organizados por período e por atividade. Para esse propósito, utilizamos imagens das listas de palavras, geradas com a ferramenta *WordList*, para, desta forma, ter uma visão geral das palavras mais recorrentes. A seguir, Figura 14, uma amostra da lista de palavras mais recorrentes da atividade “opinião” do *corpus* escrito.

Figura 14 – Palavras mais recorrentes na atividade de opinião do corpus escrito.

N	Word	Freq.	%	Texts	%
1	DE	111	5.00	13	92.86
2	QUE	111	5.00	14	100.00
3	LA	80	3.61	14	100.00
4	Y	49	2.21	12	85.71
5	EL	48	2.16	11	78.57
6	EN	40	1.80	13	92.86
7	LAS	40	1.80	14	100.00
8	ESPAÑOL	37	1.67	13	92.86
9	A	36	1.62	12	85.71
10	ES	36	1.62	12	85.71
11	UN	34	1.53	13	92.86
12	NO	32	1.44	13	92.86
13	LOS	31	1.40	12	85.71
14	UNA	30	1.35	12	85.71
15	LENGUA	28	1.26	11	78.57
16	ME	26	1.17	11	78.57
17	CON	24	1.08	9	64.29
18	MÁS	24	1.08	10	71.43
19	POR	24	1.08	10	71.43
20	PARA	21	0.95	10	71.43

frequency alphabetical statistics filenames notes

660 entries Row 8 ESPAÑOL

Fonte: a autora.

Além da visão geral das palavras mais recorrentes, detalhamos mais a informação criando tabelas que apresentassem as dez palavras lexicais mais recorrentes. A seguir a tabela criada para a atividade de opinião do *corpus* escrito.

Figura 15 – Amostra da tabela com as palavras lexicais mais recorrentes.

Posição	Palavra	Categoría gramatical	F
1°	español	Adjetivo/substantivo	37
2°	lengua	Substantivo	28
3°	variante	Adjetivo	20
4°	ser	Verbo/substantivo	18
5°	creo	Verbo	16
6°	españa	Substantivo	15
7°	países	Substantivo	15
8°	variantes	Substantivo	14
9°	también	Advérbio	13
10°	gusta	Verbo	12

Fonte: a autora.

Na Figura 15, vemos as 10 palavras lexicais mais recorrentes, organizando as palavras pela frequência das ocorrências e incluindo a categoria gramatical de cada uma delas.

Após realizada a análise por palavras mais recorrentes, o passo seguinte foi a análise dos erros quantitativos. Para esse propósito, os dados foram agrupados por períodos, tipos de erro, número de ocorrências e representados numa tabela, criando desta forma uma visão geral. Para facilitar a leitura e interpretação dos dados obtidos, criamos gráficos que organizariam a informação por *corpus*, tipo de erro, amostrando qual era a quantidade destes em cada um dos períodos e, também, gráficos para representar a quantidade de erros em cada tipo de atividade. A seguinte imagem, Figura 16, corresponde a um dos gráficos construídos para esse propósito.

Figura 16 – Amostra do gráfico de análise quantitativa.



Fonte: a autora.

A parte final da análise de dados, correspondeu à análise qualitativa. Para esse propósito foi preciso utilizar a ferramenta *Concord* do programa *WordSmith Tools 6.0*, desta forma, geramos as linhas de concordância que nos permitiu apreciar o lugar exato no documento onde encontra-se o “erro” e, assim, acessar o contexto da oração para poder realizar uma análise do erro mais exata.

Figura 17 – Amostra das linhas de concordância.

Concordance

N	Concordance	Set
27	<LEX [películas]> -películas- túdo <MOR [todo]> que pára <SIN [todo	B
28	con mis pádrés en fería libre y también <MOR [también]> hágó párté del pibiyí	B
29	[entiendo]> en tódos lugáres son asín <MOR [así]> / cuánto <SIN [sobre]> la	B
30	A: estóy bem <MOR [bien]> também <MOR [también]> A: soy \ nuéva	B
31	com <MOR [con]> las drógas / entáo <MOR [entonces]> ácho <LEX [creo]>	B
32	/ vño legalizár lá e vño trazér pára açá <MOR : SIN [vayan a legalizar allá y	V
33	acaba con el narcotráfico?]> D: Entáo <MOR [entonces]>, eu ácho <SIN :	B
34	<MOR [estructura]> D: Entáo <MOR [entonces]> B: Péro que hacér	B
35	[entonces]> B: Péro que hacér entáo <MOR [entonces]> D: Não <MOR [no]>	B
36	[vidas]> con múchias cósas no só <MOR [solo]> com <MOR [con]> las	B

concordance collocates plot patterns clusters timeline filenames source text notes

284 entries Row 1 s A: ¿Acá en brasíl tódas as <MO

Fonte: a autora.

Como tratava-se de uma quantidade de erros muito elevada, tomamos somente cinco exemplos para cada tipo de “erro” por período, os quais foram selecionados considerando a sua representatividade quanto à quantidade de vezes que o erro se repetiu e, também, considerando os “erros” que poderiam ter uma mesma raiz e fossem similares a outros erros que registraram. Desta forma, criamos um formato para organizar a análise de cada erro, compondo por: erro, forma esperada e a possível causa que levou a cometer o erro.

A seguir, Figura 18, uma imagem do formato utilizado para a análise qualitativa.

Figura 18 – Amostra do formato utilizado para a análise qualitativa.

1. "...tódas as <MOR [las]> escuelas..."

Erro	Forma esperada	Possível causa
tas	las	Transferência do português.

2. "...inscribir pesonalmente <MOR [personalmente]>..."

Erro	Forma esperada	Possível causa
pesonalmente	personalmente	Mistura a forma correta do português "pessoalmente" com a forma correta do espanhol "personalmente".

3. "...se compra más <MOR [más]> marijuana..."

Erro	Forma esperada	Possível causa
más	más	Transferência do português.

4. "...-películas- tudo <MOR [todo]>..."

Erro	Forma esperada	Possível causa
tudo	todo	Transferência do português.

5. "...lugáres son asín <MOR [asi]>..."

Erro	Forma esperada	Possível causa
asiñ	asi	Transferência do português.

Fonte: a autora.

A metodologia, foi então, localizar os erros mais representativos por período, fazer a escrita do “erro”, seguido da forma correta esperada pra esse “erro”, finalizando com uma resposta para a possível causa.

O passo final da metodologia consistiu nas considerações finais, parte que destaca conclusões e pareceres sobre todo o presente trabalho de pesquisa, podendo ser encontrada no final desse texto.

4 ANÁLISE DE DADOS

4.1 Introdução

A seguinte seção, constitui-se como a mais extensa do nosso trabalho. Após o processo de compilação do *corpus* linguístico, ou seja, depois de cumprir todas as etapas de planejamento, aplicação das atividades de coleta, transcrições e digitações dos dados coletados nas atividades, identificação das idiossincrasias e etiquetagem dos “erros”, foi possível avançar até o ponto da análise de dados.

Essa análise compõe-se por diferentes subseções e tipos, as quais se apresentarão numa ordem cronológica.

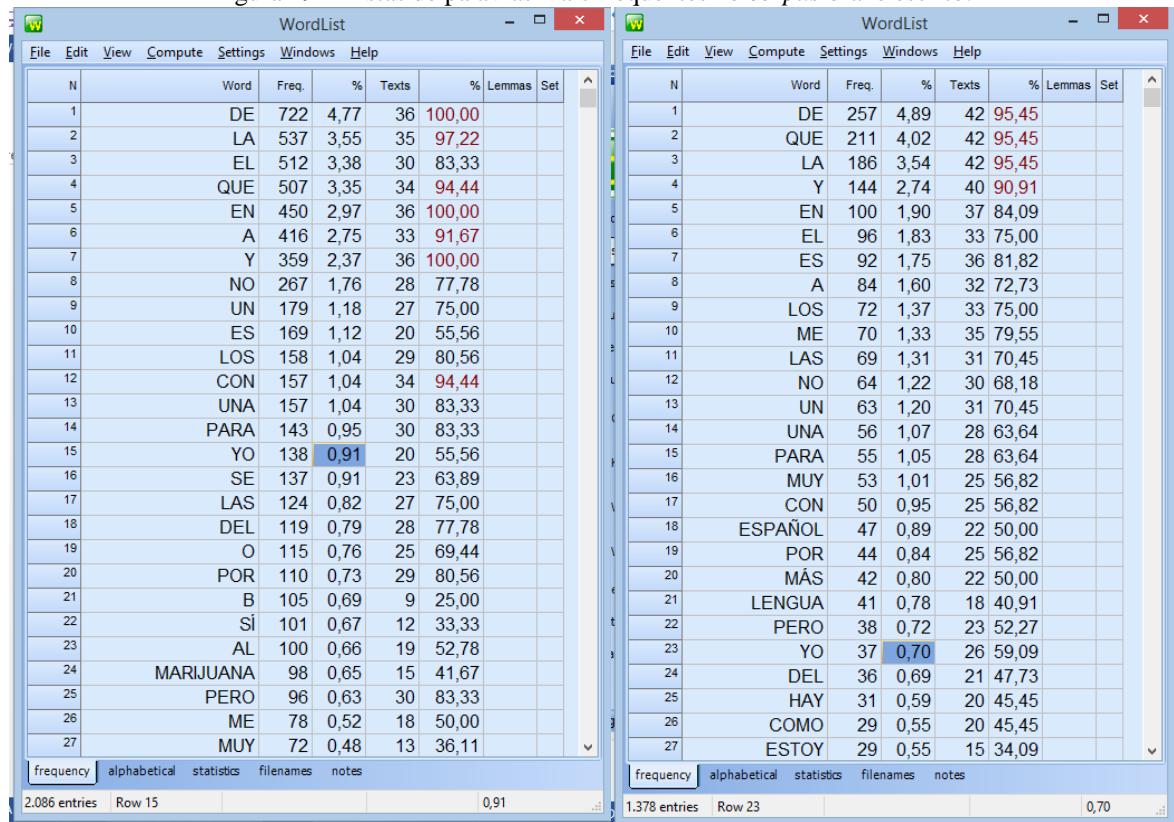
A primeira dessas subseções foi concebida com o intuito de ser apresentada no exame de qualificação, como análises preliminares, contendo apenas um recorte do *corpus*. Os dados do segundo período, com o qual também realizamos algumas considerações parciais sobre o trabalho, ainda encontrava-se em progresso, prestes ao exame de qualificação.

Para a construção das seguintes subseções, a organização do trabalho teve de continuar com o padrão cronológico, pois era prioritário determinar qual era a extensão dos *subcorpora*, para depois continuar com as posteriores análises, pois, sem esta informação prévia, seria complexo determinar quais seriam os tipos de análises pertinentes de executar. Sendo assim, as análises efetuadas giraram em torno de quais eram as palavras mais recorrentes e no desenvolvimento de uma análise quantitativa e qualitativa.

4.2 Análises Preliminares

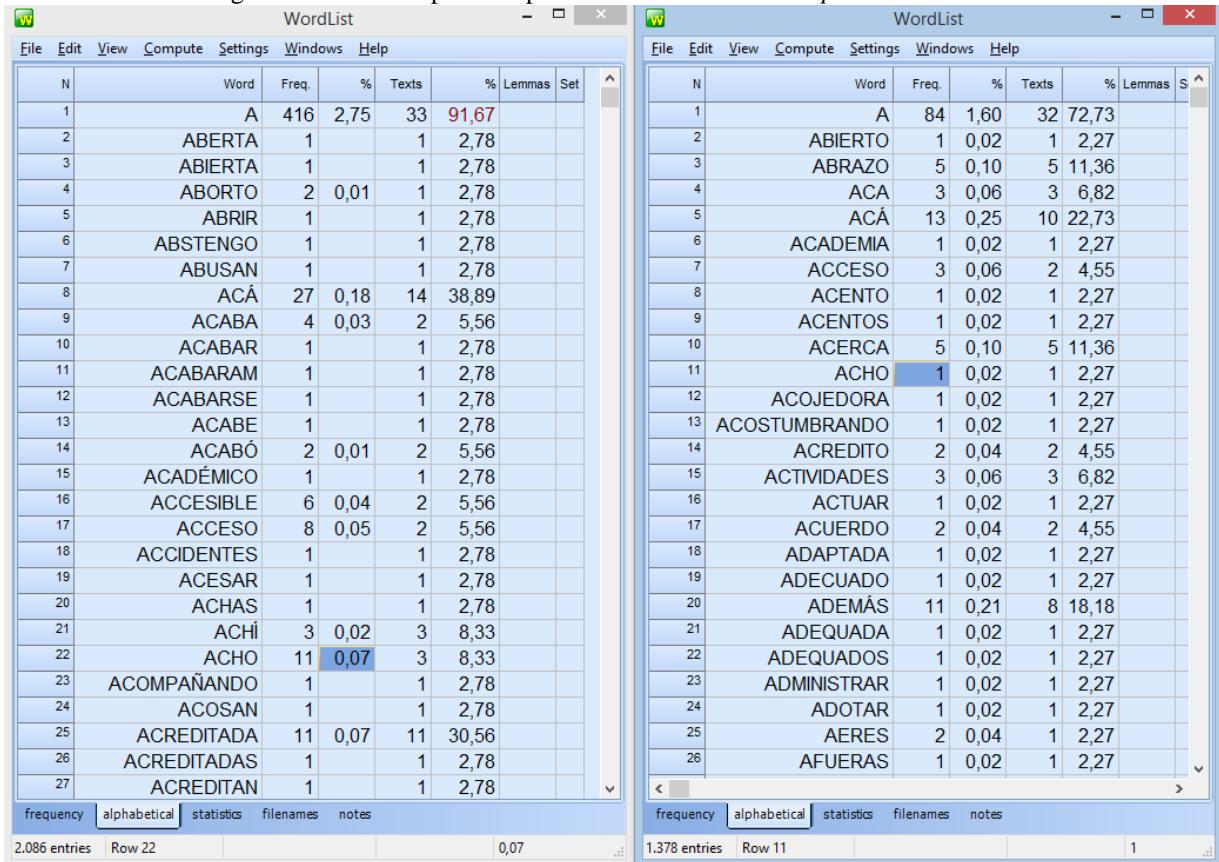
Para a etapa de exame de qualificação fez-se imprescindível saber gerenciar o programa *Word Smith Tools*, desta maneira, pudemos acessar os dados estatísticos preliminares da presente pesquisa.

Na sequência, as listas com as palavras mais frequentes possibilitam um estudo contrastivo entre os termos de ambos os *corpora*, podendo considerar a recorrência e o uso das palavras:

Figura 19 – Listas de palavras mais frequentes no *corpus* oral e escrito.

Fonte: a autora.

A seguir, nas listas de palavras organizadas por ordem alfabética, já podem ser apreciadas algumas ocorrências de interlíngua, que serão objeto de nossa análise descritiva. O procedimento seguinte será a classificação e etiquetagem dos “erros”, conforme os critérios próprios do modelo da AE.

Figura 20 – Lista de palavras por ordem alfabética no *corpus* oral e escrito.

Fonte: a autora.

Para a presente seção, decidimos fazer um recorte do *corpus* de estudo. Desta forma, tomamos como amostra para as análises preliminares as produções orais e escritas do segundo período dos alunos do Curso de Letras. Desta maneira, foram etiquetados somente os arquivos correspondentes ao período antes mencionado.

Tabela 7 – Extensão do recorte do *corpus* de estudo para as análises preliminares.

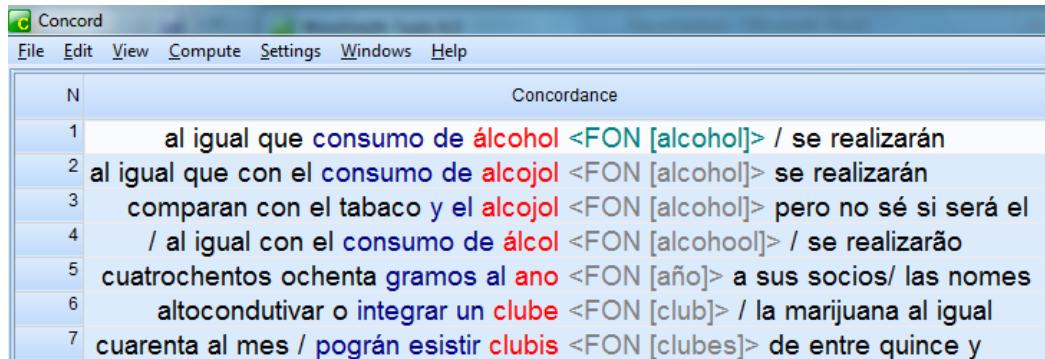
	Documentos	Itens (<i>Tokens</i>)	Formas (<i>Types</i>)	Razão Forma/Item (%)
Corpus Oral	10	3.122	921	29.50%
Corpus Escrito	11	866	396	46.32%
TOTAL	21	3.988	1.164	29.27%

Fonte: a autora.

As etiquetas utilizadas para esta etapa consideraram somente os tipos de “erros” gramaticais supracitados na seção teórica do presente texto. As etiquetas utilizadas foram: <FON> para os tipos de “erros” fonológicos; <ORT> para os tipos de “erros” ortográficos; <MOR> para os tipos de “erros” morfológicos; <SIN> para os tipos de “erros” sintáticos e;

<LEX> para os tipos de “erros” lexicais.

Figura 21 – Linhas de concordância pela categoria FON do *corpus* oral.



N	Concordance
1	al igual que consumo de álcohol <FON [alcohol]> / se realizarán
2	al igual que con el consumo de alcojol <FON [alcohol]> se realizarán
3	comparan con el tabaco y el alcojol <FON [alcohol]> pero no sé si será el
4	/ al igual con el consumo de ácol <FON [alcohol]> / se realizarão
5	cuatrocientos ochenta gramos al año <FON [año]> a sus socios/ las nomes
6	altocondutivar o integrar un club <FON [club]> / la marijuana al igual
7	cuarenta al mes / pográn esistir clubis <FON [clubes]> de entre quince y

Fonte: a autora.

Na Figura 21 é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” FON do *corpus* oral, lembrando o conceito adotado em Durão (2004), “fonológico, quando o erro afeta confusões entre fonemas”.

A organização foi alfabética, mostrando neste recorte as primeiras sete linhas de concordância. Podemos ver da esquerda para a direita: o número da linha de concordância, seguido de uma sentença que, na parte do meio, contém uma etiqueta com o tipo de “erro” FON, o qual inclui entre [] qual é a forma correta para “erro”. Do lado esquerdo da etiqueta, podemos ver, em vermelho, a palavra que contém o “erro”.

Para esta categoria, encontramos 73 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número dois:

(01) *Al igual que con el consumo de alcojol*<FON [alcohol]> *se realizarán*

Neste caso o “erro” é do tipo fonológico porque na pronúncia da palavra “alcohol” o H não é pronunciado, é mudo. No exemplo, foi pronunciado como fonema fricativo [x], que equivaleria a um J do espanhol “*alcojol*”.

Figura 22 – Linhas de concordância pela categoria MOR do *corpus* oral.

N	Concordance
127	en el mundo luego que el país en lo < MOR [el] > intento de reducir la violencia
128	como son teles? B: teóricamente lo < MOR [el] > gobierno distribui todas
129	que ten- ten- tienen la lengua materna lo < MOR [el] > español / espero conseguir
130	oficina del correo / alli se le tomaron los < MOR [las] > huellas del índice izquierdo
131	las drogas / então acho que será mais < MOR [más] > difícil en Brasil B: Sí /
132	que conduz al tráfico/ se compra mais < MOR [más] > marijuana do que cocaína
133	el decreto entre en vigencia el marteses < MOR [martes] > pero recién al fin del año
134	entre quince miembros que con un masin < MOR [máximo] > de noventa y nove
135	Me llamo masprefiro < MOR [prefiero] > que me chamem yo

Fonte: a autora.

Na Figura 22, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” MOR do *corpus* oral, lembrando o conceito adotado em Durão (2004) “morfológico, quando o ‘erro’ está relacionado com a forma”.

A organização foi alfabética, mostrando neste recorte desde a linha 127 até a 135 de concordância.

Para esta categoria, encontramos 254 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número 127:

(02): *en el mundo luego que el país en lo<MOR [el]> intento de reducir la violencia*

Trata-se de um “erro” de tipo morfológico, pois existe uma confusão na conformação do artigo masculino em espanhol, possivelmente causado por influência do artigo masculino do português “o”, que levou o falante, por intuição, a produzir um “lo”, em vez de um “el”, que corresponderia à forma correta em espanhol.

Figura 23 – Linhas de concordância pela categoria SIN do *corpus* oral.

N	Concordance
37	Brasil es un país muy major <SIN : LEX [mucho más
38	não C: Acho que para medicina <SIN [para la medicina]> no hay
39	el narcotráfico mesmo mismo <SIN [Y será que realmente
40	B: picar algo A: vamos B: acá na <SIN [en el]> lanchonete de la
41	tienen mucho o no tienen nadie <SIN [o nadie tiene]>? B: tienen
42	D: Yo también no C: É difíciu né <SIN [no es difícil?]> D: Sí / muy
43	interesada debe cocurrir a oficina <SIN [a la oficina]> del correo /
44	filmes -películas- tudo que para <SIN [todo para]> llamar la
45	já é possível D: Mas se for para <SIN : MOR [si es para]>

Fonte: a autora.

Na Figura 23, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” SIN do *corpus* oral, lembrando o conceito adotado em Durão (2004) “sintático, quando o ‘erro’ está vinculado com a ordem da construção oracional”.

A organização foi alfabética, mostrando neste recorte, desde a linha de concordância 37 até a 45.

Para esta categoria, encontramos 53 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número 43:

(03): interesada debe concurrir a oficina <SIN [a la oficina]> del correo

Trata-se de um “erro” sintático, pois vemos que o falante produz um “a”, que toma como uma interferência linguística do português para o espanhol. A forma “à” possui o sentido de “a la” em espanhol, que corresponderia à forma correta esperada para este contexto.

Figura 24 – Linhas de concordância pela categoria LEX do *corpus* oral.

N	Concordance
1	/ no ahora / por lo menos no ahora C: Acho <LEX [creo]> que la población ainda nao
2	cosas no só com las drogas / então acho <LEX [creo]> que será mais difícil en Brasil
3	para perjudicar / acho que já não C: Acho <LEX [creo]> que para medicina no hay
4	que para medicina no hay dúvida D: Acho <LEX [creo]> que por la vida B: No ha
5	como en Uruguay / yo penso así y acho <LEX [creo]> A: También Brasil es un país
6	profesora de portugués o de inglés acho <LEX [creo]> que só eso
7	B: sí sí ¿cuándo empiezas a dar aula <LEX [clases]>? A: hoy B: ¿cuándo

Fonte: a autora.

Na Figura 24, é possível apreciar um recorte das linhas de concordância

correspondentes à categoria de “erro” LEX do *corpus* oral, lembrando o conceito adotado em Durão (2004), “Léxico-semântico, quando o ‘erro’ afeta o significado das lexias”.

A organização foi alfabética, mostrando neste recorte as primeiras sete linhas de concordância.

Para esta categoria, encontramos 24 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número dois:

(04): *então acho <LEX [creo]> que será mais difícil em Brasil*

Neste caso, trata-se de um “erro” de tipo lexical, o falante toma a palavra do português “acho” que no espanhol não existe. “Creo” é a forma correta no espanhol para este contexto.

Figura 25 – Linhas de concordância pela categoria ORT do *corpus* escrito.

Concord	
	File Edit View Compute Settings Windows Help
N	Concordance
1	/ Hoy empezé las classes de español aca <ORT [acá]> en Buenos Aeres como
2	/ tengo varias cosas para contarte // Aca <ORT [acá]> en Mesico no hay muchas
3	Entré hace poco tiempo aca <ORT [acá]> en la UFU / pero hasta ahora
4	de valorizacón del español latino-americano <ORT [latinoamericano]>
5	trabajo el dobro para completarlos / así <ORT [así]> poso hacer las actividades que
6	la UFU es una boa facultad / la assignatura <ORT [asignatura]> de español es bien
7	actividades que mejor son para cada classe <ORT [clase]> // Tengo de traer peliculas /

Fonte: a autora.

Na Figura 25, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” ORT do *corpus* escrito, lembrando o conceito adotado em Durão (2004), “ortográfico, quando afeta a grafia das palavras”.

A organização foi alfabética, mostrando neste recorte as primeiras sete linhas de concordância.

Para esta categoria, encontramos 44 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número seis:

(05): *la **assignatura**<ORT [asignatura]> de español es*

Aqui o “erro” é do tipo ortográfico. O falante, possivelmente, toma como referência

a palavra em português "assinatura" e comete o “erro” de escrevê-la do mesmo modo para o espanhol. A forma correta para o espanhol só possui uma letra “s”: “asignatura”.

Figura 26 – Linhas de concordância pela categoria MOR do *corpus* escrito.

Concordance	
N	
26	sobre la dominación eropea < MOR [europea] > e la emportancia de
27	Miguel / Yo estoy muy bien / la escola < MOR [escuela] > es muy buena / los
28	los tipos y el alumno debería escolher < MOR [escoger] > o mejor para el // A
29	habito a la lengua / o que mas escuto < MOR [escucho] >
30	// Y creo que todos que aquí estan < MOR : SIN [Y creo que todos los
31	receptivas // Espero que tudo esteja < MOR [esté] > bien // Hasta luego //
32	posui sus variedades / y cada falante < MOR [hablante] > prefiere usar su
33	ver que las instituciones financeras < MOR [financieras] > como el banco
34	rebeberan a mi muy bien // Yo fue < MOR [fui] > a algunos lugares
--	

Fonte: a autora.

Na Figura 26, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” MOR do *corpus* escrito. A organização foi alfabética, mostrando neste recorte as linhas de concordância desde a 26 até a 34.

Para esta categoria, encontramos 74 ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número 26:

(06): *sobre la dominación eropea* < MOR [europea] >

Este exemplo corresponde ao tipo de “erro” morfológico, porque a forma da palavra é muito similar à forma correta em espanhol “europea”, mas o falante escreve “eropea”, omitindo o vogal “u”.

Figura 27 – Linhas de concordância pela categoria SIN do *corpus* escrito.

Concordance	
N	
1	es difundir la lengua / eso da mucho certo < SIN : MOR [resulta mucho] > / pero las
2	en el curso de Letras / yo soy enamorada < SIN [me encanta] > por la estructura física /
3	(por que treino mucho) por otro siento falta < SIN [extraño] > de lo sotaque minero / e
4	la gramatica // Hasta ahora / estoy gustando < SIN [me está gustando] > y aprendiendo
5	que treino mucho) por otro siento falta de lo < SIN : MOR [el] > sotaque minero / e todos
6	políticas és más complicado do que < SIN : AD[que] > en Brasil // El regime no és
7	poso hacer las actividades que mejor son < SIN [que son mejores] > para cada classe /

Fonte: a autora.

Na Figura 27, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” SIN do *corpus* escrito. A organização foi alfabética, mostrando neste recorte as primeiras sete linhas de concordância.

Para esta categoria, encontramos oito ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número um:

(07): *eso da mucho certo*<SIN: MOR [resulta mucho]>

No exemplo supracitado podemos encontrar um “erro” duplo, ou seja, de tipo sintático e também de tipo morfológico, porque o falante escreve misturado algumas palavras em português e outras em espanhol. A forma correta em espanhol seria "resulta mucho".

Figura 28 – Linhas de concordância pela categoria LEX do *corpus* escrito.

N	Concordance
1	<LEX [creo]> que la UFU es una boa facultad
2	assignatura de español es bien desenvolvida <LEX [desarrollada]> / ultrapassan solamente
3	en las escuelas y no solo en ensino <LEX [enseñanza]> de lenguas / pero
4	Acredito que la UFU es una boa facultad <LEX [universidad]> / la assignatura de
5	un organismo fija las normas más tambiem há <LEX [existe]> la question politica y cultural en
6	// Como vos sabés yo estoy acá lecionando <LEX [enseñando]> el Portugués y
7	en la enseñanza médica portanto tengo muy <LEX [muchos]> alumnos / pero no parecen

Fonte: a autora.

Na Figura 28, é possível apreciar um recorte das linhas de concordância correspondentes à categoria de “erro” LEX do *corpus* escrito. A organização foi alfabética, mostrando neste recorte as primeiras sete linhas de concordância.

Para esta categoria, encontramos nove ocorrências, mas para o presente caso apresentarei só um exemplo de análise de “erro”, tomando como referência a linha de concordância número dois:

(08): *asignatura de español es bien desenvolvida*<LEX [desarrollada]>

O tipo de “erro” neste caso é lexical, pois o falante escreve uma palavra trazida do português “desenvolvida”, mas em espanhol a palavra é totalmente diferente “desarrollada”.

A seguir, apresentamos os resultados dispostos em tabela, indicando a quantificação dos “erros” encontrados, a partir de cada uma das categorias propostas para a análise, tanto no

corpus oral quanto no escrito.

Tabela 8 – Ocorrências de erros em *corpus* oral e escrito.

	FON	ORT	MOR	SIN	LEX	TOTAL
<i>Corpus Oral</i>	73		254	53	24	404
<i>Corpus Escrito</i>		44	74	8	9	135

Fonte: a autora.

Na Tabela 8 podemos observar que a maior quantidade de “erros” foi encontrada na categoria de “erros” morfológicos (MOR). Em ambos os *corpora*, esse foi o “erro” mais recorrente, com 254 ocorrências no *corpus* oral e 74 ocorrências no *corpus* escrito. Na sequência seguiram os “erros” fonológicos (FON), com um total de 73 ocorrências (somente no *corpus* oral) e no *corpus* escrito, por sua vez, os “erros” ortográficos (ORT), com 44 ocorrências. “Erros” sintáticos (SIN) e lexicais (LEX) foram os menos frequentes: 53 “erros” sintáticos no *corpus* oral, 8 “erros” sintáticos no *corpus* escrito; 24 “erros” lexicais no *corpus* oral e 9 “erros” no *corpus* escrito.

Esses resultados preliminares, tomados apenas nas produções escritas e orais dos participantes do segundo período, conduzem às seguintes reflexões: Os “erros” morfológicos são os mais recorrentes, possivelmente devido à grande cota de intuição e predição de como deveriam ser as diversas formas verbais, principalmente. Enquanto falantes de português como língua materna e perante uma língua tão próxima como o espanhol, os estudantes tentariam adivinhar muitas das formas ou, simplesmente, fariam predições, a partir de uma aquisição ainda primária na LE, acabando cometendo “erros” na forma dessas palavras. Por outro lado, cabe mencionar que esta análise preliminar, por tratar-se do grupo de estudantes com menos horas de estudo de espanhol, serviu de base para a análise das produções dos demais períodos.

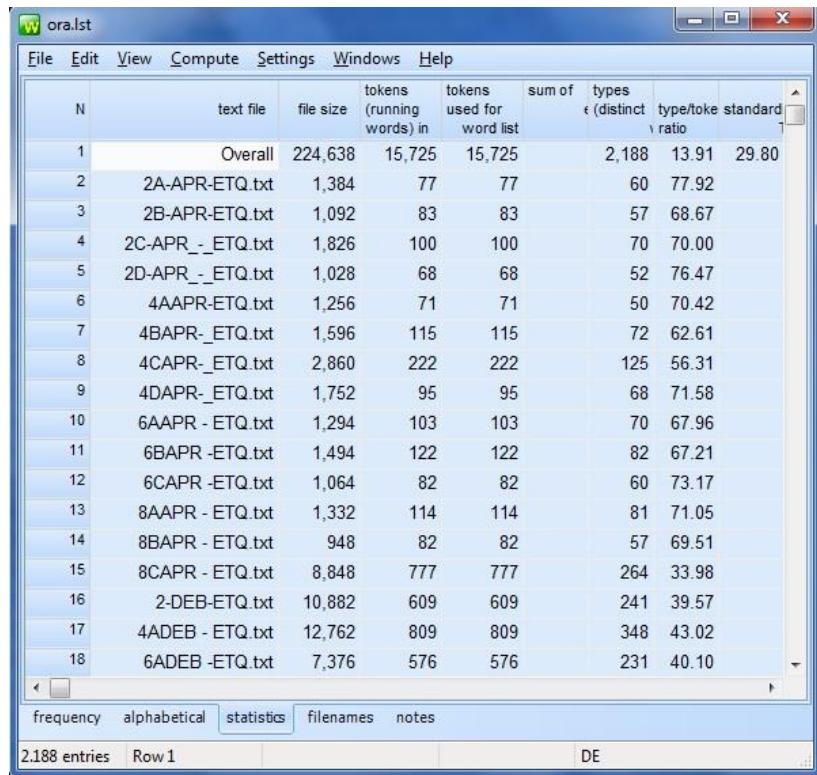
4.3 Extensão dos *subcorpora* por período

Uma das grandes tarefas no desenvolvimento da presente pesquisa foi a compilação do *corpus* de estudo. Para isso, foi preciso aplicar questionários de sondagem, selecionar os participantes, construir atividades, aplicar atividades, digitalizar, digitar e transcrever as informações para, desse modo, poder cumprir com uma das metas mais elementares da nossa pesquisa: saber a extensão do *corpus* que será analisado a partir do presente capítulo. Para começar a referirmos à dimensão do *corpus*, faremos uma divisão entre o *corpus* oral, e o *corpus*

escrito, para, desta forma, poder observar os dados por separado e, logo, organizados em uma tabela.

A seguir apresentamos a Figura 29 que apresenta os dados estatísticos mais gerais do *subcorpus* oral:

Figura 29 – Extensão do *corpus* oral.



N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/total ratio	standard
1	Overall	224,638	15,725	15,725	2,188	13.91	29.80
2	2A-APR-ETQ.txt	1,384	77	77	60	77.92	
3	2B-APR-ETQ.txt	1,092	83	83	57	68.67	
4	2C-APR_- ETQ.txt	1,826	100	100	70	70.00	
5	2D-APR_- ETQ.txt	1,028	68	68	52	76.47	
6	4AAPR-ETQ.txt	1,256	71	71	50	70.42	
7	4BAPR_- ETQ.txt	1,596	115	115	72	62.61	
8	4CAPR_- ETQ.txt	2,860	222	222	125	56.31	
9	4DAPR_- ETQ.txt	1,752	95	95	68	71.58	
10	6AAPR - ETQ.txt	1,294	103	103	70	67.96	
11	6BAPR - ETQ.txt	1,494	122	122	82	67.21	
12	6CAPR - ETQ.txt	1,064	82	82	60	73.17	
13	8AAPR - ETQ.txt	1,332	114	114	81	71.05	
14	8BAPR - ETQ.txt	948	82	82	57	69.51	
15	8CAPR - ETQ.txt	8,848	777	777	264	33.98	
16	2-DEB-ETQ.txt	10,882	609	609	241	39.57	
17	4ADEB - ETQ.txt	12,762	809	809	348	43.02	
18	6ADEB - ETQ.txt	7,376	576	576	231	40.10	

Fonte: a autora.

Dentre todos os dados apresentados na Figura 29, selecionamos os mais pertinentes à pesquisa para serem analisados a seguir.

O número total de arquivos utilizados foi de 37, incluindo todos os períodos e todas as atividades orais. O tamanho do *corpus* dimensiona-se a partir da quantidade de palavras que este possui, no caso, o número total de itens (*tokens*) foi de 15.725, dentre os quais as palavras diferentes foram somente 2.188 formas (*types*) gerando, desse modo, uma razão forma/item de 13,91%.

Figura 30 – Extensão do *corpus* escrito.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token ratio
1	Overall	72,296	5,173	5,117	1,362	26.62
2	2A-CAR-ETQ.txt	2,022	138	138	98	71.01
3	2B-CAR-ETQ.txt	830	45	43	36	83.72
4	2C-CAR-ETQ.txt	2,258	117	115	82	71.30
5	2D-CAR-ETQ.txt	1,348	74	72	64	88.89
6	4A-CAR-ETQ.txt	2,072	125	125	101	80.80
7	4B-CAR-ETQ.txt	1,842	157	155	104	67.10
8	4C-CAR-ETQ.txt	1,068	59	56	48	85.71
9	4D-CAR-ETQ.txt	1,676	101	99	74	74.75
10	6A-CAR-ETQ.txt	2,134	175	171	114	66.67
11	6B-CAR-ETQ.txt	1,112	86	86	67	77.91
12	6C-CAR-ETQ.txt	1,306	103	103	77	74.76
13	6D-CAR-ETQ.txt	1,868	140	138	101	73.19
14	8A-CAR-ETQ.txt	2,884	239	236	143	60.59
15	8B-CAR-ETQ.txt	1,442	118	116	87	75.00
16	8C-CAR-ETQ.txt	1,022	81	79	65	82.28
17	2A-DES-ETQ.txt	912	52	52	45	86.54
18	2B-DES-ETQ.txt	840	48	48	39	81.25

Fonte: a autora.

Quanto ao *corpus* escrito, diga-se que este também foi constituído considerando todos os períodos e todas as atividades de produção escrita, sendo assim, o número de arquivos utilizados foi de 43. A quantidade total de itens do *corpus* foi de 5.173 itens, sendo 1.362 o número de formas, gerando uma razão forma/item de 26,62%.

Tabela 9 – Extensão do *corpus* de estudo.

	Documentos	Itens (<i>Tokens</i>)	Formas (<i>Types</i>)	Razão Forma/Item (%)
Corpus Oral	37	15.725	2.188	13,91
Corpus Escrito	43	5.173	1.362	26,62
TOTAL	80	20.898	3.390	16,27

Fonte: a autora.

Na Tabela 9, podemos ver os dados estatísticos do *corpus* de estudo. Uma das observações principais que podemos fazer é o fato do *corpus* oral conter um tamanho maior ao do *corpus* escrito quanto ao número de itens. Mas também, aprecia-se que a razão forma/item é inferior no *corpus* oral, se comparado ao *corpus* escrito. Esse fato deriva do tamanho diferente dos *corpora*: como a extensão do *corpus* oral é aproximadamente três vezes maior que o *corpus* escrito, a possibilidade de o número das repetições aumentarem é maior também. Ou seja,

possivelmente, se o *corpus* escrito tivesse um maior número de itens, a razão forma/item tenderia a descer, como foi o caso do *corpus* oral.

Quanto ao *corpus* oral, ele foi organizado de duas maneiras, por períodos (segundo, quarto, sexto e oitavo períodos) e por tipo de atividade (apresentação, leitura, diálogo e debate). Começaremos observando os resultados estatísticos do *corpus* oral organizado por períodos.

Fonte: a autora.

Figura 31 – Amostra da extensão do *corpus* oral no segundo período.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of tokens	types (distinct)	type/oke ratio	standardde
1	Overall	52,524	3,133	3,133		921	29.40	
2	2A-APR-ETQ.txt	1,384	77	77		60	77.92	
3	2B-APR-ETQ.txt	1,092	83	83		57	68.67	
4	2C-APR_-ETQ.txt	1,826	100	100		70	70.00	
5	2D-APR_-ETQ.txt	1,028	68	68		52	76.47	
6	2-DEB-ETQ.txt	10,882	609	609		241	39.57	
7	2A-DIA-ETQ.txt	6,886	408	408		199	48.77	
8	2B-DIA-ETQ.txt	6,240	399	399		193	48.37	
9	2A-LEI-ETQ.txt	10,254	562	562		300	53.38	
10	2B-LEI-ETQ.txt	8,920	561	561		288	51.34	
11	2D-LEI-ETQ.txt	4,012	266	266		150	56.39	

Na Figura 31, podemos apreciar os seguintes dados: foram considerados para este recorte um total de dez arquivos, a quantidade de itens foi de 3.133. Em relação ao número de formas, o número chegou a 921, obtendo uma razão forma/item de 29,40%.

Observaremos a continuação os dados estatísticos correspondentes ao quarto período.

Figura 32 – Amostra dos dados estatísticos do *corpus* oral no quarto período.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of tokens	types (distinct)	type/oke ratio	standardde
1	Overall	67,260	4,451	4,451		1,059	23.79	
2	4ALEI - ETQ.txt	9,722	571	571		296	51.84	
3	4BLEI - ETQ.txt	9,010	566	566		286	50.53	
4	4CLEI-ETQ.txt	7,964	568	568		281	49.47	
5	4DLEI -_ETQ.txt	8,902	575	575		286	49.74	
6	4ADIA-ETQ.txt	6,080	476	476		191	40.13	
7	4BDIA - ETQ.txt	5,356	383	383		188	49.09	
8	4ADEB - ETQ.txt	12,762	809	809		348	43.02	
9	4AAPR-ETQ.txt	1,256	71	71		50	70.42	
10	4BAPR_-ETQ.txt	1,596	115	115		72	62.61	
11	4CAPR_-ETQ.txt	2,860	222	222		125	56.31	
12	4DAPR_-ETQ.txt	1,752	95	95		68	71.58	

Fonte: a autora.

Na Figura 32, podemos observar que se trata de um recorte do *corpus* oral correspondente ao quarto período do Curso de Letras, no qual foram utilizados onze arquivos. A quantidade de itens encontrados foi de 4.451, constituindo-se como palavras diferentes ou formas 1.059, com uma razão forma/item de 23,79%. A continuação os dados estatísticos do *Corpus Oral*, correspondem ao sexto período.

Figura 33 – Amostra dos dados estatísticos do *corpus* oral no sexto período.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token ratio
1	Overall	42,002	3.142	3.142	758	24.12
2	6AAPR - ETQ.txt	1,294	103	103	70	67.96
3	6BAPR - ETQ.txt	1,494	122	122	82	67.21
4	6CAPR - ETQ.txt	1,064	82	82	60	73.17
5	6ADEB - ETQ.txt	7,376	576	576	231	40.10
6	6ADIA - ETQ.txt	6,272	533	533	200	37.52
7	6ALEI - ETQ.txt	8,272	565	565	280	49.56
8	6BLEI - ETQ.txt	8,210	582	582	282	48.45
9	6CLEI - ETQ.txt	8,020	579	579	283	48.88

Fonte: a autora.

Na Figura 33, vemos que foram utilizados oito arquivos para poder atingir os seguintes dados: o número de itens foi de 3.142, quanto ao número de formas, este foi de 758, provendo assim uma razão forma/item de 24,12%.

Figura 34 – Amostra dos dados estatísticos do *corpus* oral no oitavo período.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token ratio
1	Overall	62,852	4,999	4,999	998	19.96
2	8ALEI - ETQ.txt	8,052	575	575	282	49.04
3	8BLEI - ETQ.txt	7,376	574	574	279	48.61
4	8CLEI - ETQ.txt	7,594	566	566	278	49.12
5	8ADIA - ETQ.txt	5,822	532	532	198	37.22
6	8ADEB - ETQ.txt	22,880	1,779	1,779	471	26.48
7	8AAPR - ETQ.txt	1,332	114	114	81	71.05
8	8BAPR - ETQ.txt	948	82	82	57	69.51
9	8CAPR - ETQ.txt	8,848	777	777	264	33.98

Fonte: a autora.

Na Figura 34, vemos que para obter os dados estatísticos do oitavo período foram utilizados um total de oito arquivos, os quais deram como resultados as seguintes informações: o número de itens foi de 4.999, as formas encontradas foram 998 e a razão forma/item foi de 19,96%.

Na Tabela 10, a seguir, apresenta-se os dados estatísticos do *corpus* oral separados por período:

Tabela 10 – Resumo dos dados estatísticos do *corpus* oral por períodos.

Período	Documentos	Itens (<i>Tokens</i>)	Formas (<i>Types</i>)	Razão Forma/Item (%)
2º	10	3.133	921	29,40
4º	11	4.451	1.059	23,79
6º	8	3.142	758	24,12
8º	8	4.999	998	19,96
TOTAL	37	15.725	2.188	13,91

Fonte: a autora.

Na Tabela 10, encontram-se organizados os dados estatísticos extraídos do *corpus* oral separado por períodos. Segundo os dados, podemos concluir o seguinte: no segundo período a produção oral foi a menor, isto pode acontecer devido ser um dos períodos iniciais do Curso de Letras, com exposição às disciplinas da especialidade de espanhol menor e, portanto, o domínio da língua ainda encontra-se em um estágio pouco desenvolvido. Já em relação ao quarto e sexto período, vemos que a produção oral aumenta no seu número de palavras, como também as horas de exposição às disciplinas da especialidade de língua espanhola. No oitavo e último período a produção oral é notoriamente maior, considerando além que o número de arquivos neste período foi somente oito, a quantidade de palavras produzidas foi acrescentada de maneira visível devido ao fato de tratar-se do período que possui uma exposição às disciplinas da especialidade de espanhol mais alta em todo o percurso do Curso de Letras.

4.4 Extensão dos *subcorpora* por atividade

Nas imagens que se seguem, podemos apreciar os dados estatísticos obtidos por atividade no *corpus* oral, começando pela atividade denominada “apresentação”:

Figura 35 – Dados estatísticos da atividade “apresentação”, recorte do *corpus* oral.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token ratio	standard
1	Overall	27,774	2,111	2,111	603	28.56	
2	2A-APR-ETQ.txt	1,384	77	77	60	77.92	
3	2B-APR-ETQ.txt	1,092	83	83	57	68.67	
4	2C-APR_-ETQ.txt	1,826	100	100	70	70.00	
5	2D-APR_-ETQ.txt	1,028	68	68	52	76.47	
6	4AAPR-ETQ.txt	1,256	71	71	50	70.42	
7	4BAPR_-ETQ.txt	1,596	115	115	72	62.61	
8	4CAPR_-ETQ.txt	2,860	222	222	125	56.31	
9	4DAPR_-ETQ.txt	1,752	95	95	68	71.58	
10	6AAPR - ETQ.txt	1,294	103	103	70	67.96	
11	6BAPR -ETQ.txt	1,494	122	122	82	67.21	
12	6CAPR -ETQ.txt	1,064	82	82	60	73.17	
13	8AAPR - ETQ.txt	1,332	114	114	81	71.05	
14	8BAPR - ETQ.txt	948	82	82	57	69.51	
15	8CAPR - ETQ.txt	8,848	777	777	264	33.98	

Fonte: a autora.

Na Figura 35, observamos a atividade denominada “apresentação”, que registrou os seguintes dados: foram catorze produções orais no total, com 2.111 itens, 603 formas e uma razão forma/item de 28,56%. Em relação à atividade denominada “debate” os dados estatísticos foram os seguintes:

Figura 36 – Dados estatísticos da atividade “debate”, recorte do *corpus* oral.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token ratio	standard
1	Overall	53,900	3,773	3,773	923	24.46	29.80
2	2-DEB-ETQ.txt	10,882	609	609	241	39.57	
3	4ADEB - ETQ.txt	12,762	809	809	348	43.02	
4	6ADEB -ETQ.txt	7,376	576	576	231	40.10	
5	8ADEB -ETQ.txt	22,880	1,779	1,779	471	26.48	29.80

Fonte: a autora.

Na Figura 36, podemos observar os dados que foram encontrados na atividade “debate”, podemos resumi-los da seguinte maneira: utilizou-se um total de quatro arquivos, o número total de palavras ou “itens” foi de 3.773, constituindo-se como formas 923, gerando uma razão forma/item de 24,46%.

Figura 37 – Dados estatísticos da atividade “diálogo”, recorte do *corpus* oral.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token standard ratio
1	Overall	36,656	2,731	2,731	692	25.34
2	2A-DIA-ETQ.txt	6,886	408	408	199	48.77
3	2B-DIA-ETQ.txt	6,240	399	399	193	48.37
4	4ADIA-ETQ.txt	6,080	476	476	191	40.13
5	4BDIA - ETQ.txt	5,356	383	383	188	49.09
6	6ADIA - ETQ.txt	6,272	533	533	200	37.52
7	8ADIA - ETQ.txt	5,822	532	532	198	37.22

Fonte: a autora.

Na Figura 37, vemos a atividade “diálogo”, podemos observar o seguinte: número total de arquivos utilizados foi de seis, número de “itens” 2.731, as formas foram 692 e a razão forma/item de 25,34%.

Figura 38 – Dados estatísticos da atividade “leitura”, recorte do *corpus* oral.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of types (distinct)	type/token standard ratio
1	Overall	106,308	7,110	7,110	644	9.06
2	2A-LEI-ETQ.txt	10,254	562	562	300	53.38
3	2B-LEI-ETQ.txt	8,920	561	561	288	51.34
4	2D-LEI-ETQ.txt	4,012	266	266	150	56.39
5	4ALEI - ETQ.txt	9,722	571	571	296	51.84
6	4BLEI - ETQ.txt	9,010	566	566	286	50.53
7	4CLEI-ETQ.txt	7,964	568	568	281	49.47
8	4DLEI - ETQ.txt	8,902	575	575	286	49.74
9	6ALEI - ETQ.txt	8,272	565	565	280	49.56
10	6BLEI - ETQ.txt	8,210	582	582	282	48.45
11	6CLEI - ETQ.txt	8,020	579	579	283	48.88
12	8ALEI - ETQ.txt	8,052	575	575	282	49.04
13	8BLEI - ETQ.txt	7,376	574	574	279	48.61
14	8CLEI - ETQ.txt	7,594	566	566	278	49.12

Fonte: a autora.

Na Figura 38, vemos os dados estatísticos encontrados no recorte correspondente à atividade “leitura”: treze arquivos foram utilizados, 7.110 o número de itens, 644 as palavras diferentes ou “formas”, originando uma razão forma/item de 9,06%.

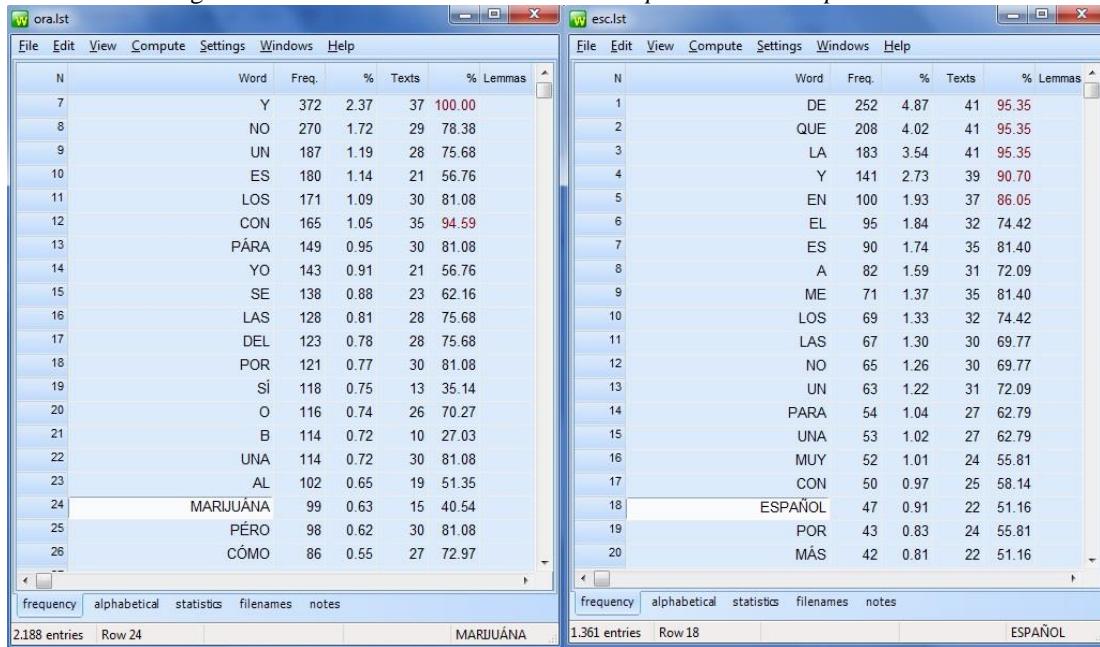
Tabela 11 – Resumo dos dados estatísticos do *corpus* oral por atividades.

Atividades	Documentos	Itens (<i>Tokens</i>)	Formas (<i>Types</i>)	Razão Forma/Item (%)
Apresentação	14	2.111	603	28.56
Debate	4	3.773	923	24.46
Diálogo	6	2.731	692	25.34
Leitura	13	7.110	644	9.06
TOTAL	37	15.725	2.188	13.91

Podemos observar na Tabela 11, os dados resumidos do *corpus* oral separado por atividades. A tabela nos revela que a atividade com maior produção de palavras ou itens foram as leituras, devido ao propósito da atividade em que cada participante lesse um mesmo texto com um determinado número de palavras. Depois, a segunda atividade com maior produção foi o debate com 3.773 itens, que além de possuir um alto número de produção, contém o maior número de palavras diferentes ou “formas”, alcançando 923. Em destaque, a apresentação pelo seu baixo número de itens produzido, com tão só 2.111 itens, o que pode dever-se à dificuldade apresentada por parte dos participantes quando se trata de falar de si próprio.

4.5 Palavras mais recorrentes no *corpus* oral e escrito

Quanto às palavras mais recorrentes nos *corpora* oral e escrito, a seguir as imagens com os dados que indicam quais foram as mais frequentes:

Figura 39 – Palavras mais recorrentes do *corpus* oral e do *corpus* escrito.

Fonte: a autora.

Em relação à Figura 39, podemos observar que as palavras lexicais mais recorrentes foram, no *corpus* oral “marijuana”, quanto ao *corpus* escrito foi “español”. Nessa, não é possível observar detalhadamente quais foram as palavras lexicais, por esse motivo, a seguir, a Tabela 12 ilustra quais foram as dez palavras com mais frequência no *corpus* oral e no *corpus* escrito. Para este fim, consideramos palavras lexicais, advérbios, e algumas conjunções.

Tabela 12 – Palavras lexicais mais recorrentes no *corpus* oral.

Posição	Palavra	Categoría gramatical	F
1°	marijuana	Substantivo	99
2°	muy	Advérbio	86
3°	venta	Substantivo	69
4°	personas	Substantivo	67
5°	entonces	Advérbio	63
6°	español	Adjetivo/substantivo	62
7°	podrá	Verbo	61
8°	consumo	Substantivo	60
9°	persona	Substantivo	60
10°	creo	Verbo	59

Fonte: a autora.

Na Tabela 12 observamos quais foram as palavras mais recorrentes do *corpus* oral.

No caso, as dez palavras mais recorrentes foram: marijuana, muy, venta, personas, entonces, español, podrá, consumo, persona e creo. Quanto às categorias gramaticais destas palavras, pode-se apreciar a predominância dos substantivos com seis ocorrências, seguido pelos advérbios com três ocorrências e verbos com duas ocorrências.

Tabela 13 – Palavras lexicais mais recorrentes no *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1°	español	Adjetivo	47
2°	lengua	Substantivo	39
3°	hay	Verbo	30
4°	estoy	Verbo	29
5°	mucho	Advérbio	28
6°	son	Verbo	28
7°	también	Advérbio	27
8°	curso	Substantivo	26
9°	creo	Verbo	24
10°	alunos	Substantivo	21

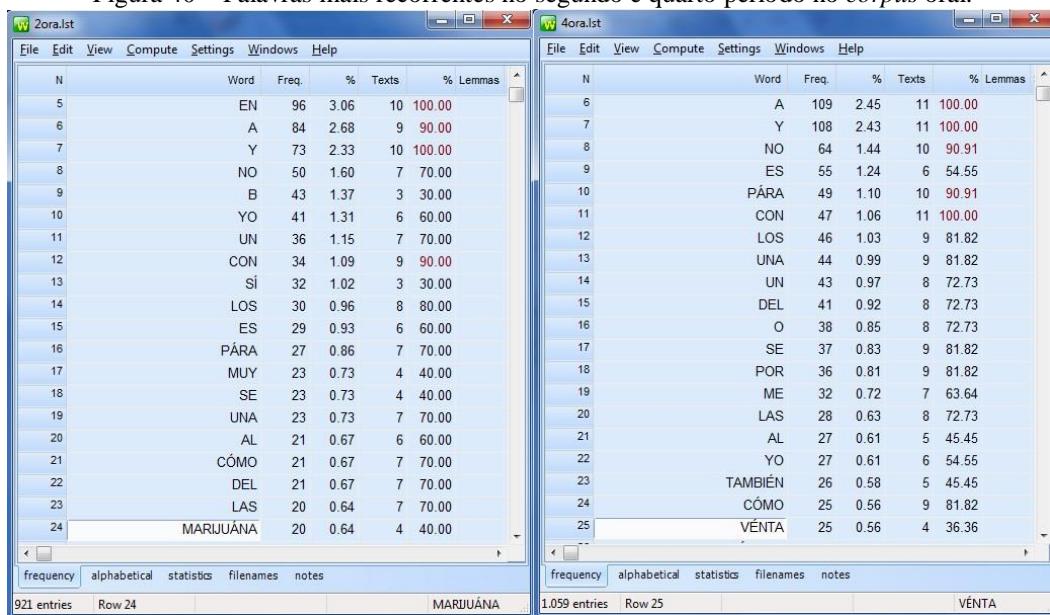
Fonte: a autora.

Na Tabela 13, podemos apreciar quais são as dez palavras mais recorrentes no *corpus* escrito. Nesta oportunidade, foram: español, lengua, hay, estoy, mucho, son, también, curso, creo e alunos. Quanto às categorias gramáticas dessas palavras, é possível afirmar que a predominância foi: verbos com quatro ocorrências, substantivos com três, seguidos pelos advérbios com duas ocorrências e, finalmente, os adjetivos com apenas uma ocorrência.

4.5.1 Palavras mais recorrentes no *corpus oral* por período

A respeito das palavras mais frequentes, a seguir, podemos observar quais foram em cada um dos períodos.

Figura 40 – Palavras mais recorrentes no segundo e quarto período no *corpus oral*.



Fonte: a autora.

Na Figura 40, vemos à esquerda os resultados da lista de todas as palavras do *corpus* oral fazendo um recorte no segundo período e, à direita, o recorte correspondente ao quarto período.

Tabela 14 – Palavras lexicais mais recorrentes do segundo período no *corpus* oral.

Posição	Palavra	Categoría grammatical	F
1º	Muy	Advérbio	23
2º	marijuana	Substantivo	20
3º	creo	Verbo	18
4º	español	Adjetivo	17
5º	son	Verbo	14
6º	consumo	Substantivo	13
7º	también	Advérbio	13
8º	acho	Verbo	11
9º	comprar	Verbo	11
10º	entonces	Advérbio	11

Fonte: a autora.

Na Tabela 14, podemos observar quais foram as palavras lexicais mais utilizadas no segundo período do *corpus* oral. As dez palavras mais recorrentes foram: *muy*, *marijuana*, *creo*, *español*, *son*, *consumo*, *también*, *acho*, *comprar* e *entonces*. Quanto às categorias gramaticais mais utilizadas, estas foram os verbos, com quatro ocorrências, advérbios com três, dois substantivos e um adjetivo.

Tabela 15 – Palavras lexicais mais recorrentes do quarto período no *corpus* oral.

Posição	Palavra	Categoría gramatical	F
1º	también	Advérbio	26
2º	venta	Substantivo	25
3º	entonces	Advérbio	23
4º	español	Adjetivo	23
5º	marijuana	Substantivo	20
6º	muy	Advérbio	20
7º	podrá	Verbo	19
8º	creo	Verbo	18
9º	comprar	Verbo	17
10º	está	Verbo	17

Fonte: a autora.

Na Tabela 15, é possível apreciar quais foram as palavras lexicais mais recorrentes no quarto período do *corpus* oral. As dez palavras mais utilizadas foram: *también*, *venta*, *entonces*, *español*, *marijuana*, *muy*, *podrá*, *creo*, *comprar* e *está*. Quanto às categorias gramaticais mais utilizadas, estas foram: verbos com quatro ocorrências, advérbios três, dois substantivos e um adjetivo.

Figura 41 – Palavras mais recorrentes no sexto e oitavo período no *corpus* oral.

Fonte: a autora.

Na Figura 41, podemos apreciar, na parte esquerda, as palavras mais recorrentes no sexto período e, à direita, as mais recorrentes no oitavo período. Somente na palavra número 19 e 22 respetivamente encontramos as primeiras palavras lexicais.

Tabela 16 – Palavras lexicais mais recorrentes do sexto período no *corpus* oral.

Posição	Palavra	Categoría gramatical	F
1º	marijuana	Substantivo	25
2º	persona	Substantivo	17
3º	venta	Substantivo	17
4º	podrá	Verbo	15
5º	mucho	Advérbio	14
6º	consumo	Substantivo	12
7º	podrán	Verbo	12
8º	comprar	Verbo	11
9º	entonces	Advérbio	11
10º	farmácias	Substantivo	11

Fonte: a autora.

Na Tabela 16, observamos quais foram as palavras lexicais mais recorrentes no sexto período do *corpus* oral. Em ordem de frequência, essas foram as primeiras dez: marijuana, persona, venta, podrá, mucho, consumo, podrán, comprar, entonces, farmácias. Quanto às

categorias gramaticais, observamos que a maior quantidade de ocorrências aconteceram nas seguintes categorias: advérbios com cinco ocorrências, verbos com três, e advérbios com duas.

Tabela 17 – Palavras lexicais mais recorrentes do oitavo período no *corpus* oral.

Posição	Palavra	Categoria gramatical	F
1°	muy	Advérbio	39
2°	personas	Substantivo	36
3°	marijuana	Substantivo	34
4°	va	Verbo	31
5°	hay	Verbo	23
6°	pienso	Verbo	22
7°	consumo	Substantivo	21
8°	está	Verbo	20
9°	creo	Verbo	19
10°	venta	Substantivo	19

Fonte: a autora.

Na Tabela 17, podemos apreciar quais foram as palavras lexicais mais recorrentes. As primeiras dez palavras lexicais em ordem de frequência foram: muy, personas, marijuana, va, hay, pienso, consumo, está, creo e venta. Quanto às categorias gramaticais, as mais frequentes foram: verbos com cinco ocorrências, substantivos com quatro ocorrências e, finalmente, advérbios com uma ocorrência.

Tabela 18 – Palavras lexicais mais recorrentes por período no *corpus* oral.

Período	Palavras mais recorrentes
2º	Muy (23), marijuana (20), creo (18)
4º	Também (26), venta (25), entonces (23)
6º	Marijuana (25), persona (17), venta (17)
8º	Muy (39), personas (36), marijuana (34)

Fonte: a autora.

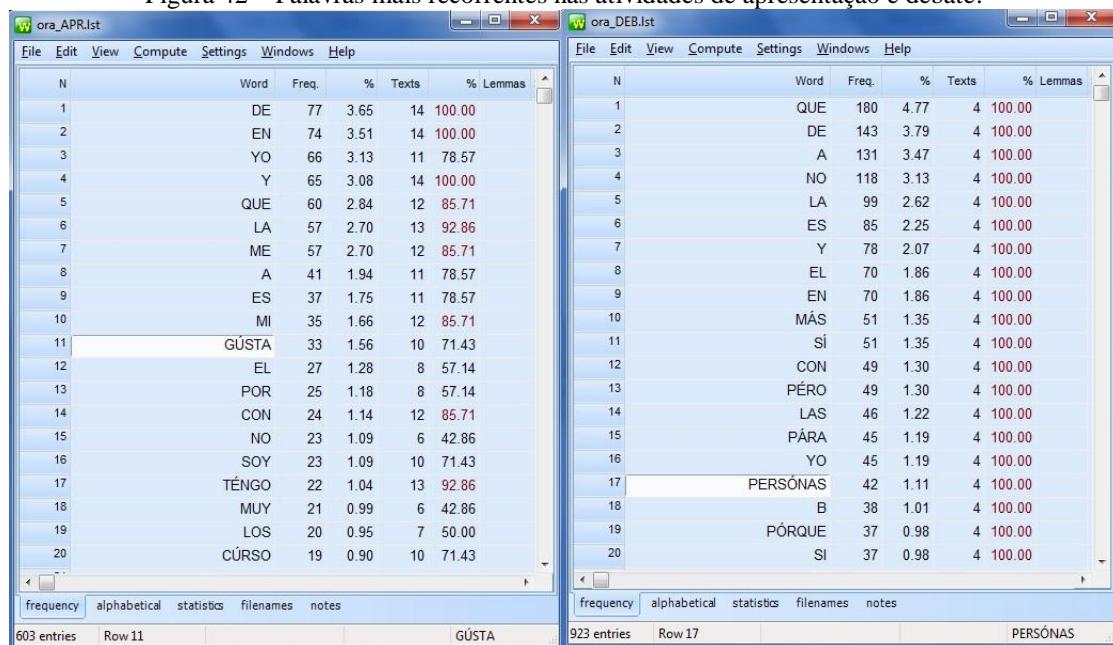
Na Tabela 18, podemos observar um resumo das informações contidas nas figuras anteriores relativas às palavras mais recorrentes por período do *corpus* oral. Nela, podemos apreciar que a palavra “marijuana” foi uma constante, e parece nas primeiras três posições do segundo, sexto e oitavo período. Outra palavra com grande frequência foi o advérbio “muy”, que se apresenta nas primeiras três posições no segundo período e no oitavo. Também a palavra

“venta” aparece nas primeiras três posições do quarto e do sexto período. Uma reflexão pertinente de fazer perante esses resultados, é que o léxico com mais frequência está relacionado com o tema da atividade proposta, no caso sobre a legalização da maconha.

4.5.2 Palavras mais recorrentes no *corpus oral* por atividade

Continuando com as palavras mais recorrentes, a seguir os recortes das atividades de apresentação, debate, diálogo e leitura correspondente ao *corpus oral*.

Figura 42 – Palavras mais recorrentes nas atividades de apresentação e debate.



Fonte: a autora.

Na Figura 42, à esquerda podemos apreciar a lista de palavras mais frequentes da atividade “apresentação” do *corpus oral*, quanto ao lado direito, podemos ver o resultado para a atividade “debate”.

Tabela 19 – Palavras lexicais mais recorrentes nas atividade de “apresentação”.

Posição	Palavra	Categoría gramatical	F
1º	gusta	Verbo	33
2º	soy	Verbo	23
3º	tengo	Verbo	22
4º	muy	Advérbio	21
5º	curso	Substantivo	19
6º	letras	Substantivo	19
7º	años	Substantivo	18
8º	mucho	Advérbio	18
9º	también	Advérbio	16
10º	universidad	Substantivo	14

Fonte: a autora.

Na Tabela 19, podemos apreciar quais foram as palavras lexicais mais frequentes na atividade “apresentação” do *corpus* oral. As dez palavras mais frequentes foram: gusta, soy, tengo, muy, curso, letras, años, mucho, también e universidad. Quanto à frequência das categorias gramaticais, as seguintes categorias foram as mais recorrentes: quatro substantivos, três verbos e três advérbios.

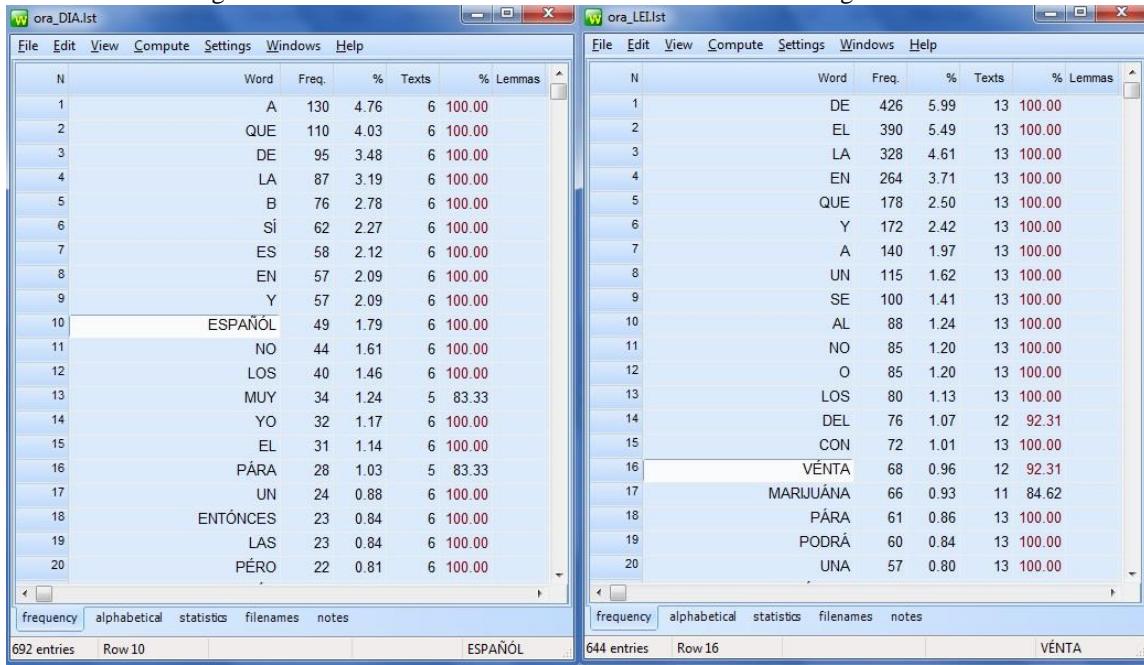
Tabela 20 – Palavras lexicais mais recorrentes nas atividade de “debate”.

Posição	Palavra	Categoría gramatical	F
1º	personas	Substantivo	42
2º	marijuana	Substantivo	33
3º	creo	Verbo	31
4º	muy	Advérbio	31
5º	va	Verbo	28
6º	también	Advérbio	27
7º	entonces	Adverbio	25
8º	pienso	Verbo	23
9º	hay	Verbo	21
10º	Brasil	Substantivo	17

Fonte: a autora.

Podemos apreciar na Tabela 20 quais foram as palavras lexicais mais recorrentes na atividade “debate” do *corpus* oral. As dez mais frequentes foram: personas, marijuana, creo, muy, va, también, entonces, pienso, hay e Brasil. Quanto às categorias gramaticais mais utilizadas, estas foram: quatro verbos, três substantivos e três advérbios.

Figura 43 – Palavras mais recorrentes nas atividades de diálogo e leitura.



Fonte: a autora.

Quanto à Figura 43, podemos ver, na parte esquerda da imagem, os dados correspondentes à atividade de diálogo, na qual a palavra que teve mais frequência foi “español”, repetindo-se 49 vezes. Na parte direita da imagem encontra-se a atividade de leitura, em que podemos apreciar a palavra lexical mais recorrente foi “venta”, com um total de 68 repetições.

Tabela 21 – Palavras lexicais mais recorrentes nas atividade de “diálogo”.

Posição	Palavra	Categoría gramatical	F
1º	español	Substantivo	49
2º	muy	Advérbio	34
3º	entonces	Advérbio	23
4º	creo	Verbo	21
5º	son	Verbo	20
6º	acá	Advérbio	19
7º	tenemos	Verbo	18
8º	hay	Verbo	17
9º	también	Advérbio	15
10º	escuela	Substantivo	13

Fonte: a autora.

Na Tabela 21, podemos observar quais foram as palavras lexicais mais recorrentes na atividade “diálogo” do *corpus* oral. As dez palavras mais frequentes foram: *español*, *muy*, *entonces*, *creo*, *son*, *acá*, *tenemos*, *hay*, *también* e *escuela*. Quanto às categorias gramaticais mais utilizadas, podemos observar que, das palavras antes mencionadas, quatro correspondem a verbos, quatro a advérbios e dois a substantivos.

Tabela 22 – Palavras lexicais mais recorrentes nas atividades de “leitura”.

Posição	Palavra	Categoría gramatical	F
1º	venta	Substantivo	68
2º	marijuana	Substantivo	66
3º	podrá	Verbo	60
4º	comprar	Verbo	50
5º	consumo	Substantivo	48
6º	persona	Substantivo	47
7º	farmácias	Substantivo	46
8º	podrán	Verbo	42
9º	público	Adjetivo	36
10º	canabis	Substantivo	33

Fonte: a autora.

Na Tabela 22, podemos apreciar quais foram as dez palavras lexicais mais frequentes na atividade “leitura” do *corpus* oral. Em ordem de maior para menor frequência, foram: *venta*, *marijuana*, *podrá*, *comprar*, *consumo*, *persona*, *farmácias*, *podrán*, *público* e *canabis*. Quanto às formas gramaticais que mais predominaram, seis formas foram substantivos, três verbos e um adjetivo.

Tabela 23 – Palavras mais recorrentes por atividade no *corpus* oral.

Atividade	Palavra mais recorrente
Apresentação	<i>gusta</i> (33), <i>soy</i> (23), <i>tengo</i> (22)
Debate	<i>personas</i> (42), <i>marijuana</i> (33), <i>creo</i> (31)
Diálogo	<i>español</i> (49), <i>muy</i> (34), <i>entonces</i> (23)
Leitura	<i>venta</i> (68), <i>marijuana</i> (66), <i>podrá</i> (60)

Fonte: a autora.

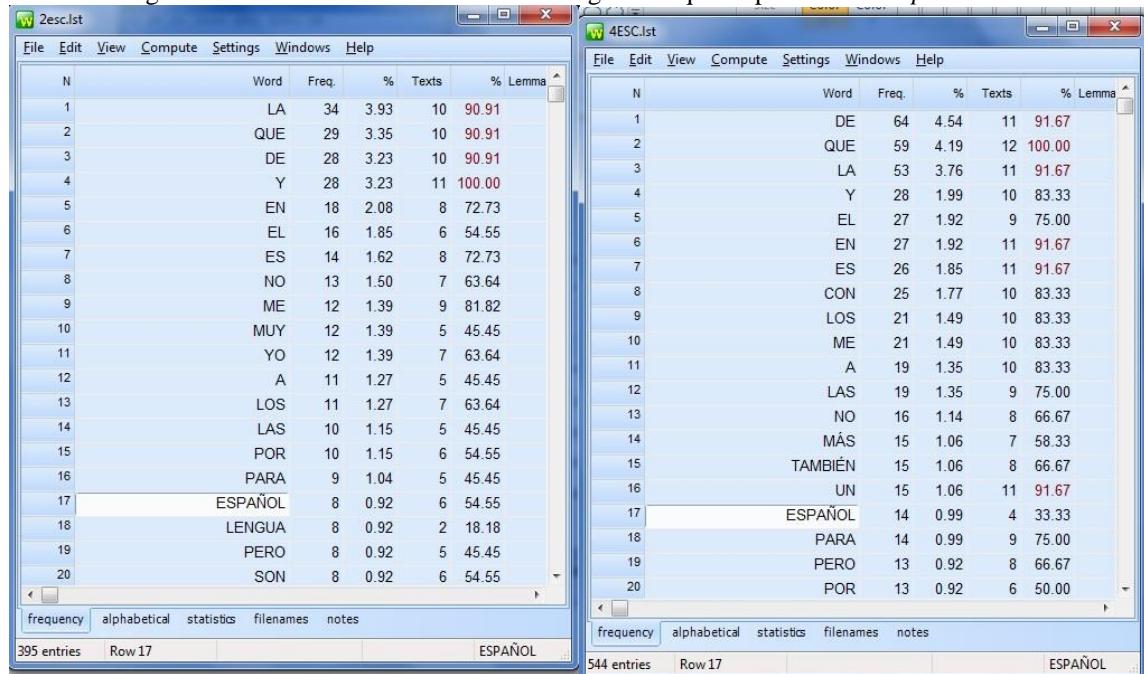
Na Tabela 23, é interessante observar como muda o léxico escolhido pelos participantes dependendo da temática da atividade proposta. Na atividade de apresentação, pediu-se para os participantes para falarem sobre si próprios e as palavras lexicais que mais

utilizaram foram os verbos “gustar”, “ser” e “tener”. Quanto à atividade “debate”, a temática proposta foi a “legalização da maconha no Uruguai”, neste caso, as palavras lexicais mais utilizadas foram “personas”, “marijuana” e “creo”. Em relação à atividade “diálogo”, os participantes tiveram uma conversação entre eles onde simulavam a chegada de um novo professor de espanhol em uma escola. Para essa atividade, as palavras lexicais mais utilizadas foram “español”, “muy” e “entonces”. Quanto à atividade “leitura”, os participantes leram um texto jornalístico sobre a legalização da maconha no Uruguai e a regulamentação do consumo, nesta oportunidades as palavras lexicais mais frequentes foram “venta”, “marijuana” e “podrá”. O tipo de léxico utilizado dependerá sempre do tipo de atividade que foi proposta.

4.5.3 Palavras mais recorrentes no *corpus* escrito por período

Em relação ao *corpus* escrito, apresentamos, a seguir, recortes dos *corpora* mencionados com o intuito de separá-los por períodos e observar quais as palavras mais frequentes.

Figura 44 – Palavras mais recorrentes no segundo e quarto período no *corpus* escrito.



Fonte: a autora.

Na Figura 44, podemos apreciar na parte esquerda da imagem o recorte correspondente para o segundo período, quanto à parte direita da imagem, amostra o recorte correspondente ao quarto período do *corpus* escrito. Em ambos, apresentam-se todas as palavras

em ordem de frequência de maior para a menor. Em destaque, a primeira palavra lexical encontrada em cada recorte.

Tabela 24 – Palavras lexicais mais recorrentes no segundo período do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1°	muy	Advérbio	12
2°	español	Substantivo	8
3°	lengua	Substantivo	8
4°	son	Verbo	8
5°	mucho	Advérbio	7
6°	estoy	Verbo	7
7°	hay	Verbo	6
8°	ser	Verbo	6
9°	bien	Advérbio	5
10°	creo	Verbo	5

Fonte: a autora.

Na Tabela 24, podemos observar as palavras lexicais mais frequentes no segundo período. As dez palavras mais frequentes, em ordem de maior a menor, foram as seguintes: muy, español, lengua, son, mucho, estoy, hay, ser, bien e creo. Quanto às categorias lexicais mais utilizadas, predominam os verbos com cinco, seguido pelos advérbios com três e dois substantivos.

Tabela 25 – Palavras lexicais mais recorrentes no quarto período do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1°	también	Advérbio	15
2°	español	Substantivo	14
3°	lengua	Substantivo	12
4°	muy	Advérbio	12
5°	estoy	Verbo	11
6°	gusta	Verbo	10
7°	curso	Substantivo	9
8°	España	Substantivo	9
9°	son	Verbo	9
10°	tiene	Verbo	8

Fonte: a autora.

À respeito da Tabela 25, podemos apreciar quais foram as palavras lexicais mais frequentes no quarto período do *corpus* escrito. As dez palavras mais recorrentes foram:

también, español, lengua, muy, estoy, gusta, curso, España, son e tiene. Quanto às categorias gramaticais, das dez palavras lexicais anteriormente citadas, quatro formas foram verbos, quatro substantivos e dois advérbios.

Figura 45 – Palavras mais recorrentes no sexto e oitavo período no *corpus* escrito.

N	Word	Freq.	%	Texts	% Lemmas
4	Y	40	2.60	10	90.91
5	EL	27	1.75	10	90.91
6	EN	26	1.69	10	90.91
7	ES	26	1.69	7	63.64
8	A	23	1.49	9	81.82
9	LOS	22	1.43	9	81.82
10	ME	22	1.43	9	81.82
11	UNA	22	1.43	9	81.82
12	LAS	19	1.23	10	90.91
13	MUY	18	1.17	8	72.73
14	PARA	18	1.17	8	72.73
15	NO	17	1.10	8	72.73
16	UN	17	1.10	8	72.73
17	#	15	0.97	7	63.64
18	MÁS	14	0.91	7	63.64
19	CON	13	0.84	8	72.73
20	HAY	12	0.78	6	54.55
21	POR	12	0.78	6	54.55
22	COMO	11	0.71	7	63.64
23	ESPAÑOL	11	0.71	6	54.55

N	Word	Freq.	%	Texts	% Lemmas
1	DE	81	5.96	9	100.00
2	QUE	56	4.12	9	100.00
3	Y	45	3.31	8	88.89
4	LA	41	3.02	9	100.00
5	A	29	2.14	7	77.78
6	EN	29	2.14	8	88.89
7	EL	25	1.84	7	77.78
8	ES	24	1.77	9	100.00
9	UN	23	1.69	7	77.78
10	LAS	19	1.40	6	66.67
11	NO	19	1.40	7	77.78
12	UNA	17	1.25	8	88.89
13	ME	16	1.18	7	77.78
14	LOS	15	1.10	6	66.67
15	#	14	1.03	7	77.78
16	ESPAÑOL	14	1.03	6	66.67
17	PARA	13	0.96	5	55.56
18	LENGUA	11	0.81	5	55.56
19	CURSO	10	0.74	3	33.33
20	DEL	10	0.74	4	44.44

Fonte: a autora.

Na parte esquerda da Figura 45, encontra-se o recorte correspondente ao sexto período, na parte direita o recorte correspondente ao oitavo período. Em ambos, amostram-se as palavras mais recorrentes em geral e, no caso, a palavra mais recorrente nos dois recortes, foi “español”, com onze e catorze repetições respectivamente.

Tabela 26 – Palavras lexicais mais recorrentes no sexto período do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1º	muy	Advérbio	18
2º	hay	Verbo	12
3º	español	Adjetivo	11
4º	creo	Verbo	10
5º	mucho	Advérbio	10
6º	alumnos	Substantivo	8
7º	estoy	Verbo	8
8º	lengua	Substantivo	8
9º	países	Substantivo	8
10º	son	Verbo	8

Fonte: a autora.

Quanto à Tabela 26, podemos observar quais foram as palavras lexicais mais recorrentes no sexto período do *corpus* escrito. Em ordem de maior para menor frequência, essas foram as dez primeiras: muy, hay, español, creo, mucho, alumnos, estoy, lengua, países e son. Quanto às categorias gramaticais que predominaram nos exemplos anteriores, podemos ver que a liderança da frequência foi dos verbos com quatro formas, seguido dos substantivos com três, dois advérbios e um adjetivo.

Tabela 27 – Palavras lexicais mais recorrentes no oitavo período do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1º	español	Adjetivo	14
2º	lengua	Substantivo	11
3º	curso	Substantivo	10
4º	muy	Advérbio	10
5º	todo	Adjetivo	9
6º	variante	Adjetivo	9
7º	alumnos	Substantivo	8
8º	hay	Verbo	8
9º	pienso	Verbo	8
10º	clases	Substantivo	7

Fonte: a autora.

Na Tabela 27, podemos observar quais foram as palavras lexicais mais recorrentes no oitavo período do *corpus* escrito. Nesta ocasião, as dez palavras mais frequentes em ordem de maior para a menor foram as seguintes: español, lengua, curso, muy, todo, variante, alumnos, hay, pienso e clases. Quanto às categorias gramaticais mais utilizadas, podemos ver que nas dez

palavras anteriormente citadas, quatro são substantivos, três adjetivos, dois verbos e um advérbio.

Tabela 28 – Palavras mais recorrentes por período no *corpus* escrito.

Período	Palavra mais recorrente
2°	español (19), muy (12), lengua (8)
4°	también (15), español (14), lengua (12)
6°	muy (18), hay (12), español (11)
8°	español (14), lengua (11), curso (10)

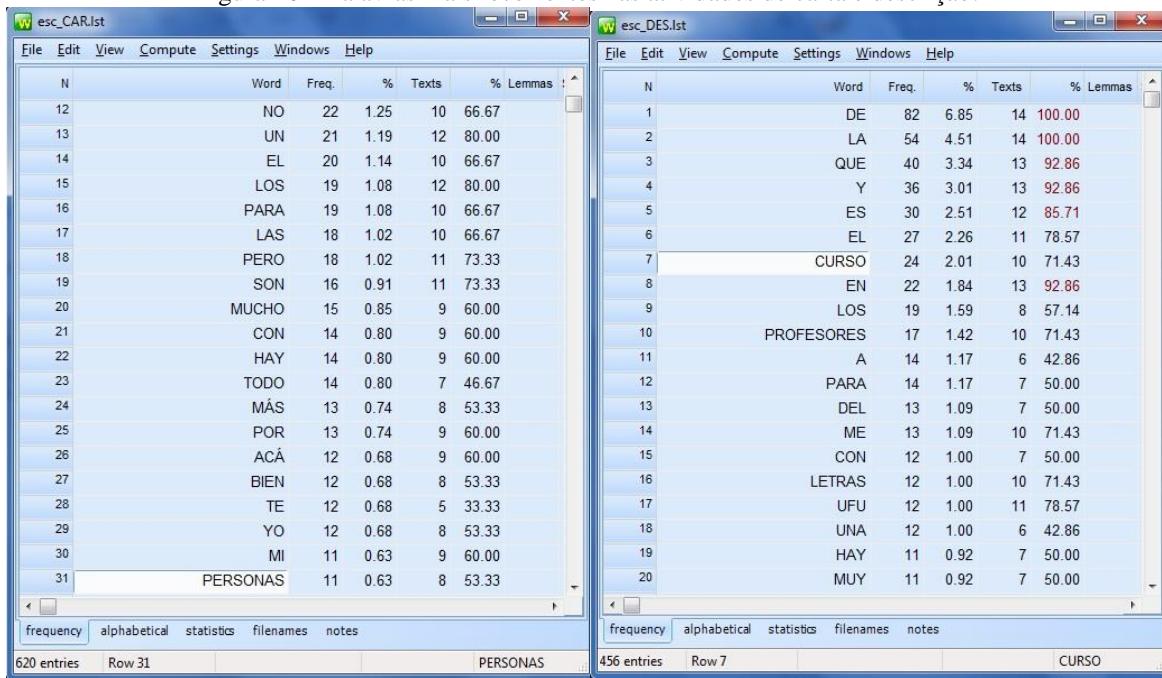
Fonte: a autora.

Na Tabela 28, podemos observar quais foram as três palavras mais recorrentes em todos os períodos do *corpus* escrito. Vemos que, no segundo período, as palavras mais utilizadas foram: español, muy e lengua. No quarto período: también, español e lengua. No sexto período as mais utilizadas foram: muy, hay e español. Quanto ao oitavo período: español, lengua e curso. O léxico utilizado em todos os períodos mantém um eixo temático que se estabelece, dessa maneira, devido ao tipo de atividades propostas para a produção escrita, a qual visava temas e reflexões à raiz da língua espanhola.

4.5.4 *Palavras mais recorrentes no corpus escrito por atividade*

Quanto ao *corpus* escrito, os dados correspondentes à frequência das palavras também foram agrupadas por atividades.

Figura 46 – Palavras mais recorrentes nas atividades de carta e descrição.



Fonte: a autora.

Na Figura 46, podemos ver, ao lado esquerdo da imagem, o recorte correspondente à atividade “carta” e, ao lado direito, o recorte corresponde à atividade “descrição”, ambas do *corpus* escrito. Foram consideradas todas as palavras, sem excluir nenhuma. Assim, podemos ver que aparecem muitas que correspondem às categorias gramaticais como preposições, conjunções e artigos.

Tabela 29 – Palavras lexicais mais recorrentes na atividade “carta” do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1º	muy	Advérbio	34
2º	estoy	Verbo	26
3º	son	Verbo	16
4º	mucho	Advérbio	15
5º	hay	Verbo	14
6º	todo	Adjetivo	14
7º	bien	Adjetivo	12
8º	personas	Substantivo	11
9º	alumnos	Substantivo	10
10º	está	Verbo	9

Fonte: a autora.

Na Tabela 29, podemos observar as palavras lexicais mais recorrentes na atividade “carta” do *corpus* escrito. Em ordem de maior a menor frequência: *muy*, *estoy*, *son*, *mucho*, *hay*, *todo*, *bien*, *personas*, *alumnos* e *está*. À respeito das categorias gramaticais mais recorrentes, tomando em conta dez palavras anteriores, podemos apreciar uma predominância dos verbos com quatro formas, dois substantivos, dois advérbios e dois adjetivos.

Tabela 30 – Palavras lexicais mais recorrentes na atividade “descrição” do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1º	curso	Substantivo	24
2º	profesores	Substantivo	17
3º	letras	Substantivo	12
4º	ufu	Substantivo	12
5º	hay	Verbo	11
6º	muy	Advérbio	11
7º	lengua	Substantivo	10
8º	también	Advérbio	9
9º	alumnos	Substantivo	8
10º	estructura	Substantivo	8

Fonte: a autora.

Na Tabela 30, podemos observar quais foram as palavras lexicais mais recorrentes para a atividade denominada “descrição” do *corpus* escrito. Nesta ocasião, as palavras mais recorrentes foram: *curso*, *profesores*, *letras*, *ufu*, *hay*, *muy*, *lengua*, *también*, *alumnos* e *estructura*. Quanto às formas gramaticais mais recorrentes, tomando em conta as dez palavras citadas anteriormente, podemos dizer que a maior frequência percebeu-se com os substantivos com sete ocorrências, seguido pelos advérbios com duas e verbo com uma.

Figura 47 – Palavras mais recorrentes na atividade de opinião.

N	Word	Freq.	%	Texts	%
1	DE	111	5.00	13	92.86
2	QUE	111	5.00	14	100.00
3	LA	80	3.61	14	100.00
4	Y	49	2.21	12	85.71
5	EL	48	2.16	11	78.57
6	EN	40	1.80	13	92.86
7	LAS	40	1.80	14	100.00
8	ESPAÑOL	37	1.67	13	92.86
9	A	36	1.62	12	85.71
10	ES	36	1.62	12	85.71
11	UN	34	1.53	13	92.86
12	NO	32	1.44	13	92.86
13	LOS	31	1.40	12	85.71
14	UNA	30	1.35	12	85.71
15	LENGUA	28	1.26	11	78.57
16	ME	26	1.17	11	78.57
17	CON	24	1.08	9	64.29
18	MÁS	24	1.08	10	71.43
19	POR	24	1.08	10	71.43
20	PARA	21	0.95	10	71.43

Fonte: a autora.

Na Figura 47, podemos observar a lista de palavras correspondentes à atividade de opinião. Nela, incluem-se todas as palavras, aparecendo a primeira palavra lexical só na posição oitava, sendo esta “español”, com um total de 37 repetições.

Tabela 31 – Palavras lexicais mais recorrentes na atividade “opinião” do *corpus* escrito.

Posição	Palavra	Categoría gramatical	F
1º	español	Adjetivo/substantivo	37
2º	lengua	Substantivo	28
3º	variante	Adjetivo	20
4º	ser	Verbo/substantivo	18
5º	creo	Verbo	16
6º	España	Substantivo	15
7º	países	Substantivo	15
8º	variantes	Substantivo	14
9º	también	Advérbio	13
10º	gusta	Verbo	12

Fonte: a autora.

Na Tabela 31, podemos observar quais foram as palavras lexicais mais recorrentes na atividade “opinião” do *corpus* escrito. Em ordem de maior a menor frequência essas foram

as palavras mais utilizadas: *español*, *lengua*, *variante*, *ser*, *creo*, *España*, *países*, *variantes*, *también* e *gusta*. Quanto às categorias gramaticais mais frequentes nas palavras anteriormente citadas, predominam os substantivos com seis ocorrências, seguido pelos verbos com três ocorrências, dois adjetivos e um advérbio.

Tabela 32 – Palavras mais recorrentes por atividade no *corpus* escrito.

Atividade	Palavra mais recorrente
Carta	<i>muy</i> (34), <i>estoy</i> (26), <i>son</i> (16)
Descrição	<i>curso</i> (24), <i>profesores</i> (17), <i>letras</i> (12)
Opinião	<i>español</i> (37), <i>lengua</i> (28), <i>variante</i> (20)

Fonte: a autora.

Na Tabela 32, podemos ver quais foram as palavras mais recorrentes por atividades no *corpus* escrito. Novamente podemos apreciar que as palavras lexicais que predominam em cada uma das atividades estão influenciadas pelo tema da atividade proposta. Por exemplo, na atividade “carta”, deviam escrever uma missiva para algum amigo ou familiar desde o estrangeiro. Na atividade de descrição, o objeto a descrever foi o Curso de Letras da UFU. Quanto à atividade “opinião” deviam refletir sobre o espanhol imposto como padrão no ensino da língua.

Quanto a presente sessão de palavras mais recorrentes, é importante mencionar que o intuito principal foi dar um panorama geral das palavras mais utilizadas, sem que necessariamente essas palavras fossem erros, ou seja, o intuito principal foi fornecer uma listagem que oferece-se a tendência real dos participantes pela utilização de certas palavras, separando a listagem por atividades e por períodos.

4.5.5 Quantitativo dos “erros” por períodos no *corpus* oral

Em relação à quantidade de “erros”, a seguir apresentam-se recortes por períodos das linhas de concordância do *corpus* oral, separados também pelos quatro tipos de “erros” para cada um dos períodos, ou seja, “erros” morfológicos, fonológicos, léxicos e sintáticos.

A seguir uma amostra dos resultados das linhas de concordância que foram considerados para a confecção das tabelas e gráficos.

Figura 48 – Linhas de concordância do segundo período para o tipo de “erro” fonológico.

N	Concordance	Set
1	[marihuana]> y sí el narcotráfico qui <FON [que]> incluso está agregado	C
2	<MOR : FON [universidad]> federáu <FON [federal]> de überlányia <FON	J
3	áños con ciudadanía uruguaya legáu <FON [legal]> o naturál y con	J
4	y cinco miémbros que con másimo <FON [máximo]> de novénta y nueve	J
5	podrá tener una concentración majór <FON [mayor]> mayór que quince por	J
6	[producto]> no puéde estar espuésto <FON [expuesto]> a la vista del	J
7	en las farmáncias habilitádas// estoh <FON [estos]> son los principales	J
8	<MOR : FON [registren]> hehístran <FON [registren]> / no será obligatório	J
9	a la persona quedará heyistrádo <FON [registrado]> el algoritmo en la	J
10	reberán <MOR [deberán]> hevistrárse <FON [registrarsel]> en una báse de	J

Fonte: a autora.

Na Figura 48, vemos o recorte do segundo período do *corpus* oral para o tipo de “erro” fonológico, no qual consta que o número de ocorrências do “erro” foi de 72.

Figura 49 – Linhas de concordância da atividade “debate” para o tipo de erro léxico.

N	Concordance	Set
1	por ejémplo / entónces ciertamente o <LEX [el]> negocio con a marijuána	A
2	creo que si / si va a querér más que lo <LEX [el]> límite / va a los traficantes	A
3	C: Yo creo que el gobiérno tiene muy <LEX [mucho]> interés en eso por	B
4	la legalización de la marijuána eu nao <LEX [no]> estóy de acuérdo pórque	B
5	También Brasil es un país muy majór <LEX : SIN [mucho más grande]>,	B
6	y será que bem regulamentáda mésimo <LEX [realmente]> esa ley D: Sí, yo	B
7	inprudéntes que utilizan álcol mismo <LEX [incluso]> aun con la prohibición	B
8	con traficantes por ejémplo / além <LEX [además]> / amás sou a favór	B
9	fui asaltáda un tiémpo atrás acá piéto <LEX [cerca]> del restoránte	B
10	el álcol es / v la marijuána no es tao <LEX [tanto]> más pernicióso aue	B

Fonte: a autora.

Na Figura 49, podemos apreciar o recorte da atividade “debate” do *corpus* oral, para o tipo de erro léxico, recorte no qual encontrou-se sessenta ocorrências do erro antes mencionado.

Contudo, organizamos a informação correspondente às ocorrências de erros, e construímos tabelas que distribuíram a informação da seguinte maneira.

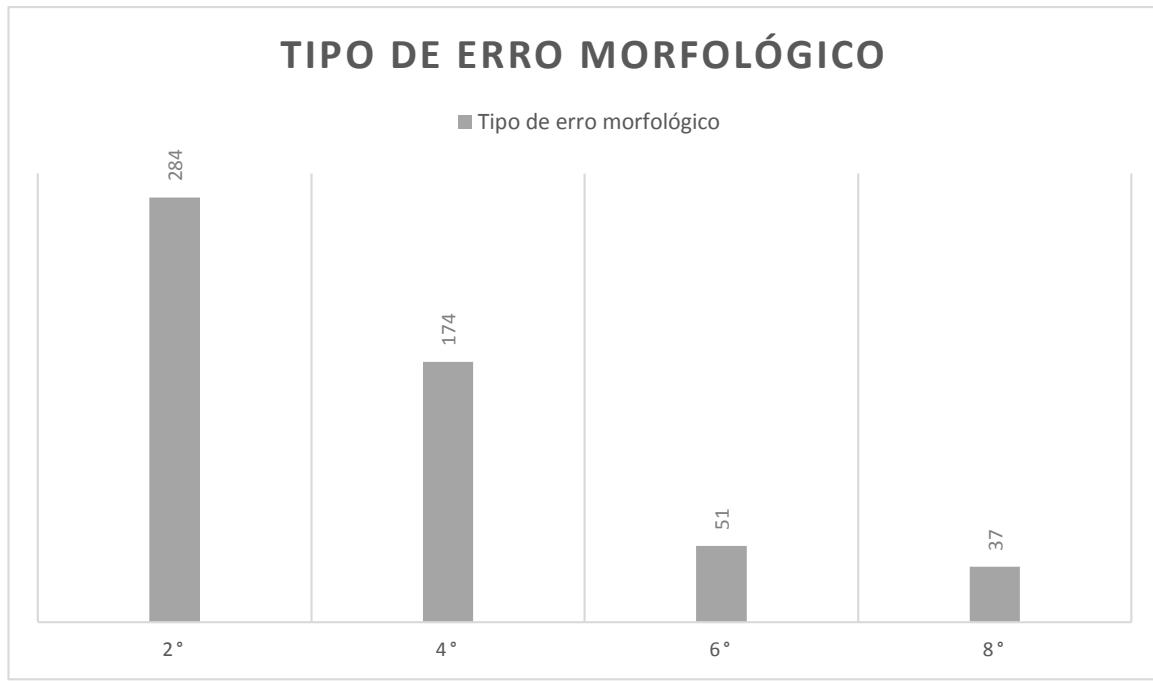
Tabela 33 – Quantitativo das ocorrências por período e por tipo de erro do *corpus* oral.

Período	Tipo de erro	O
2°	Fonológico	72
2°	Léxico	27
2°	Morfológico	284
2°	Sintático	41
4°	Fonológico	126
4°	Léxico	63
4°	Morfológico	174
4°	Sintático	28
6°	Fonológico	57
6°	Léxico	10
6°	Morfológico	51
6°	Sintático	5
8°	Fonológico	61
8°	Léxico	21
8°	Morfológico	37
8°	Sintático	8

Fonte: a autora.

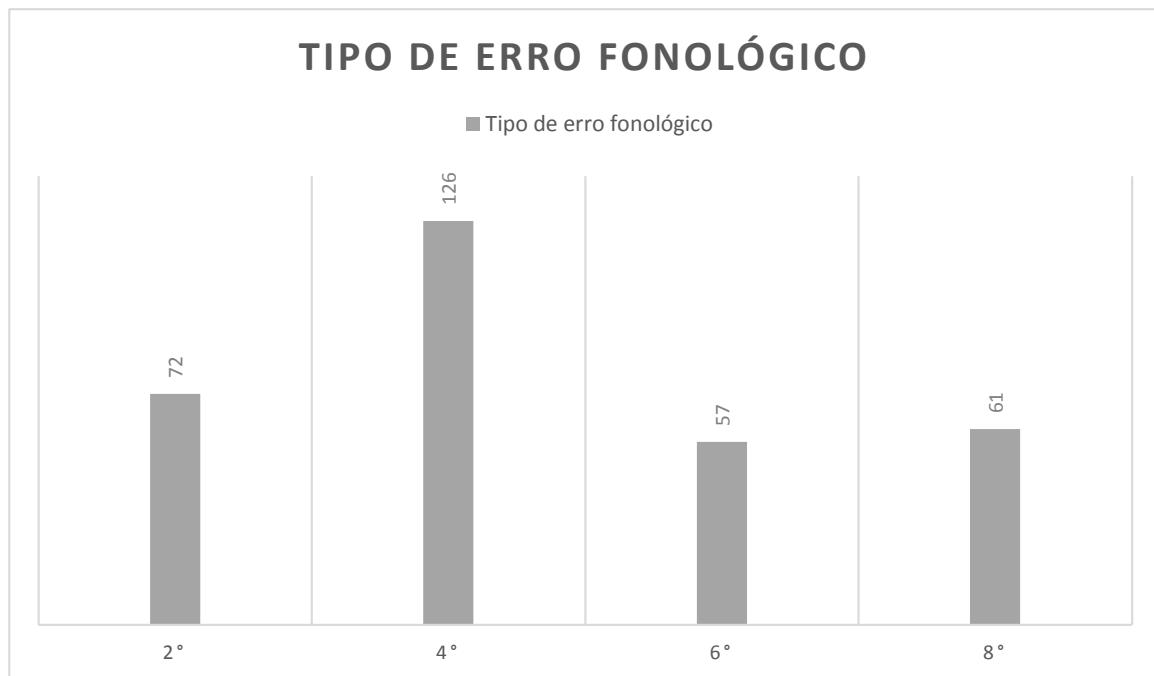
Na Tabela 33, encontram-se os dados correspondentes às ocorrências de cada um dos tipos de erros, organizados por período. Fazemos destaque no segundo período, no tipo de erro morfológico, pois contém o maior número de ocorrências, e também no sexto período no tipo de erro sintático pois possui o menor número de ocorrências com apenas 5. A seguir, apresentaremos gráficos com o intuito de fazer uma leitura mais fácil dos dados apresentados nesta tabela.

Figura 50 – Quantitativo dos erros morfológicos por período do *corpus* oral.



Fonte: a autora.

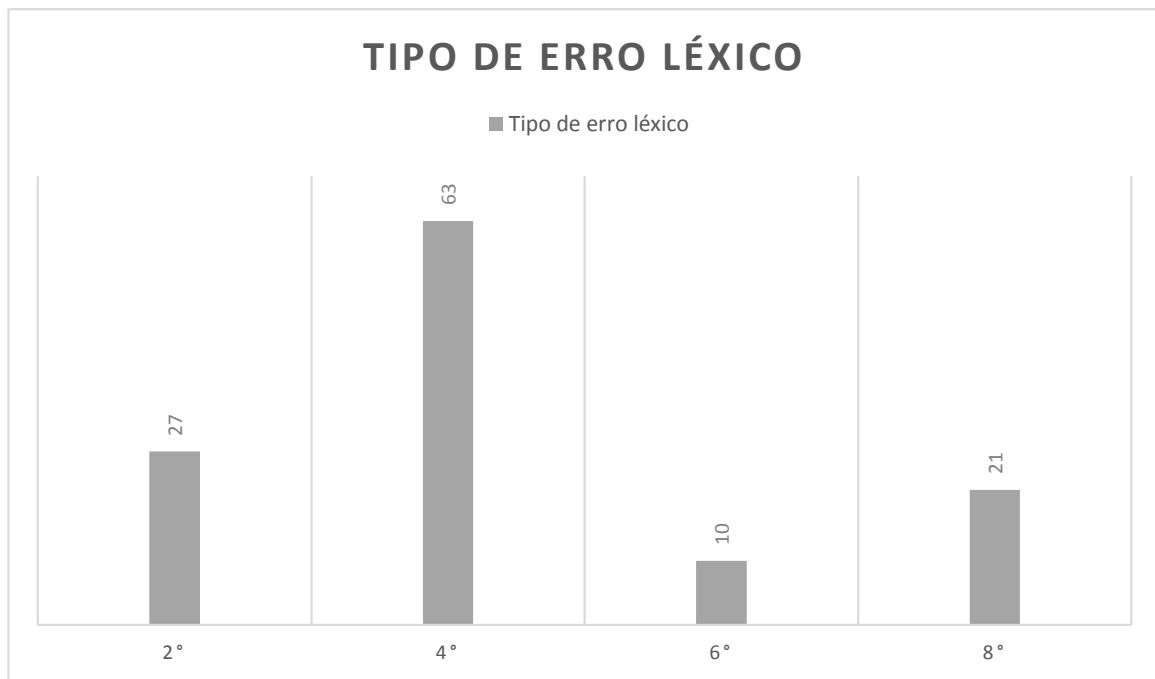
Quanto à Figura 50 e aos tipos de erros morfológicos que surgiram no *corpus* oral, é importante mencionar que se trata do tipo de erro mais copioso. Este, aparece no segundo período com um número de 284 erros, diminuindo a cada ano, dessa forma, podemos apreciar que já no quarto período o número diminui a 174, no sexto período a 51, chegando no oitavo período com apenas 37 ocorrências desse tipo de erro.

Figura 51 – Quantitativo dos erros fonológicos por período do *corpus* oral.

Fonte: a autora.

Na Figura 51, quanto ao tipo de erro fonológico, podemos apreciar que não existe uma tendência a diminuir ou aumentar o tipo de erro. De fato, é possível afirmar que o número não tem uma grande variação desde o período inicial até o final, considerando que, no segundo período, o número para este tipo de erro foi de 72 e, no último período, de 61. Cabe fazer um destaque no quarto período, pois foi o que registrou mais ocorrências desse tipo de erro (126).

Figura 52 – Quantitativo dos erros lexicais por período do *corpus* oral.



Fonte: a autora.

Na Figura 52, quanto ao tipo de erro léxico, podemos observar que não existe uma grande variação desde o período inicial até o último período, pois, no segundo período, o número de ocorrências foi de 27 e no final de 21. Novamente, convém fazer um destaque no quarto período que registrou o número mais elevado de erros desse tipo, chegando até 63 ocorrências.

Figura 53 – Quantitativo dos erros sintáticos por período do *corpus* oral.

Fonte: a autora.

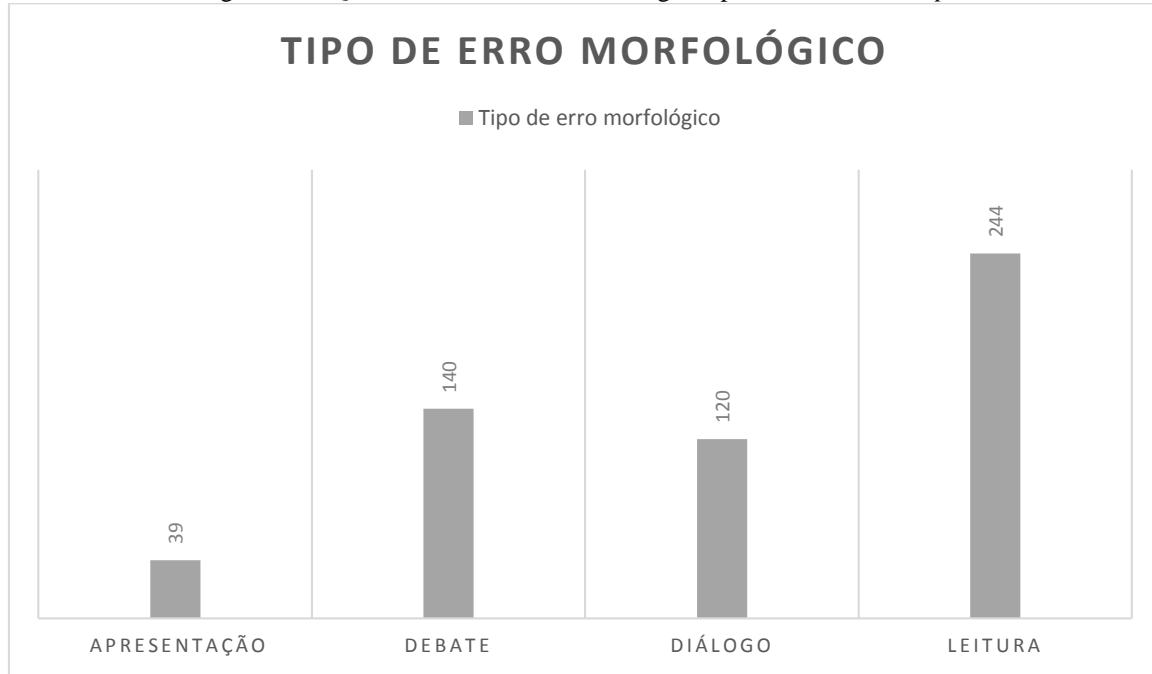
Na Figura 53, quanto ao tipo de erro sintático, podemos afirmar que existe uma tendência a diminuir no percurso do curso: os alunos do segundo período registraram 41 erros deste tipo, no quarto período o número se abrevia a 28, já, no sexto período, percebemos uma diminuição importante, pois chega-se no seu menor número, com 5 ocorrências e, no oitavo período, apreciamos que o número tem um leve aumento chegando a 8.

Tabela 34 – Quantitativo das ocorrências por atividade e por tipo de erro do *corpus* oral.

Atividade	Tipo de erro	O
Apresentação	Fonológico	21
Apresentação	Léxico	25
Apresentação	Morfológico	39
Apresentação	Sintático	22
Debate	Fonológico	68
Debate	Léxico	60
Debate	Morfológico	140
Debate	Sintático	30
Diálogo	Fonológico	7
Diálogo	Léxico	28
Diálogo	Morfológico	120
Diálogo	Sintático	26
Leitura	Fonológico	220
Leitura	Léxico	6
Leitura	Morfológico	244
Leitura	Sintático	4

Fonte: a autora.

Na Tabela 34, podemos apreciar os dados de ocorrências de erros, mas desta vez organizados por atividades. Para facilitar a leitura destes dados, apresenta-se, a seguir, a informação organizada em gráficos.

Figura 54 – Quantitativo dos erros morfológicos por atividade do *corpus* oral.

Fonte: a autora.

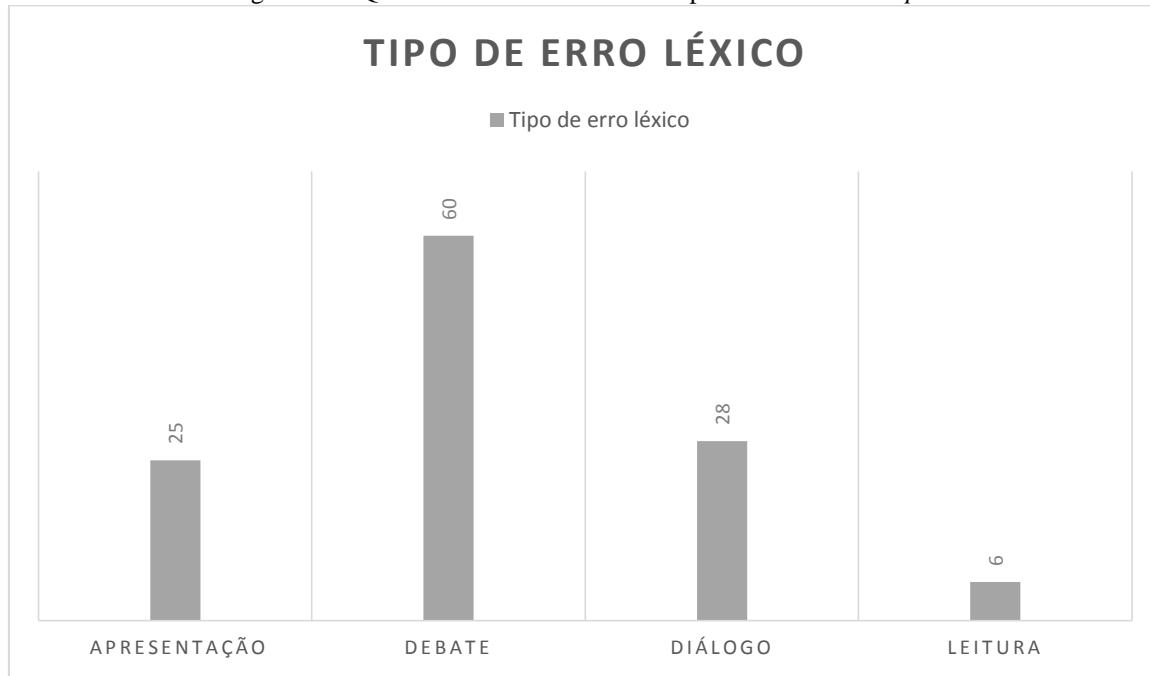
Na Figura 54, quanto ao tipo de erro morfológico, podemos apreciar que o maior número de ocorrências foi registrado na atividade de leitura com 244 ocorrências, seguido pela atividade de debate com 140, a atividade de diálogo com 120 ocorrências e, finalmente, a atividade de apresentação com 39, em que também cabe mencionar que foi a atividade com o menor número de palavras produzidas.

Figura 55 – Quantitativo dos erros fonológicos por atividade do *corpus* oral.



Fonte: a autora.

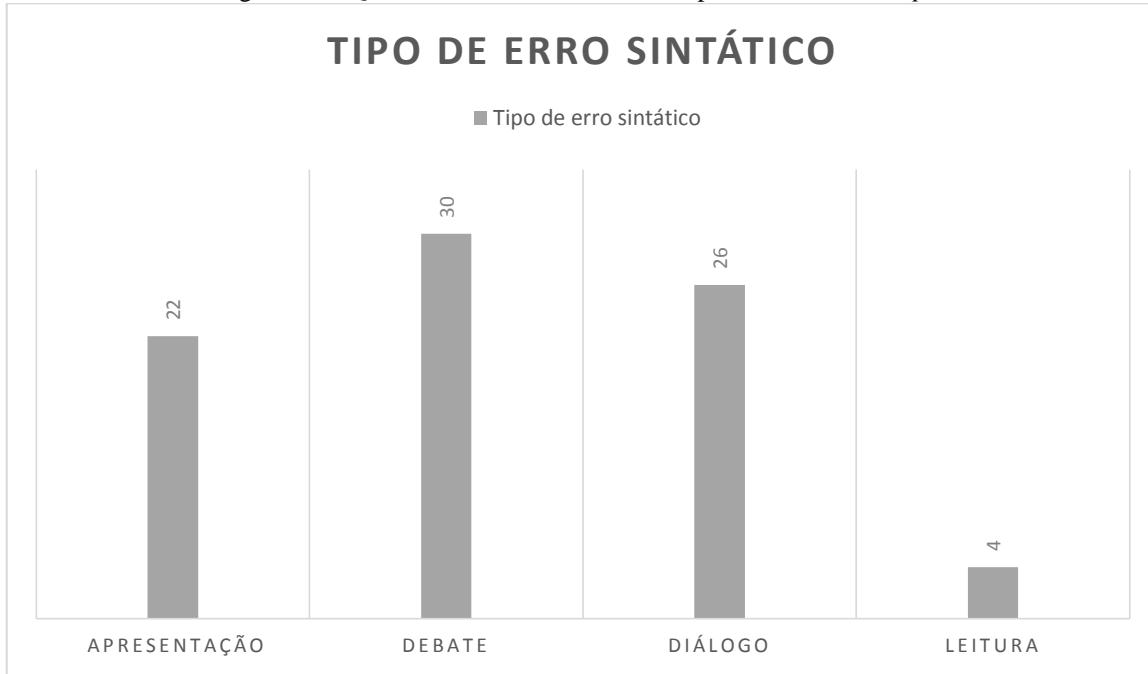
Na Figura 55, quanto ao tipo de erro fonológico, foi na atividade de leitura que se registraram 220 ocorrências, o maior número de ocorrências para este tipo de erro, seguido pela atividade de debate, em que se encontraram 68, na apresentação 21 e, finalmente, a atividade que menos ocorrências de erro fonológico registrou, foi a de diálogo, com apenas sete ocorrências.

Figura 56 – Quantitativo dos erros lexicais por atividade do *corpus* oral.

Fonte: a autora.

Na Figura 56, quanto aos erros de tipo léxico, a atividade em que mais ocorrências registraram foi o debate com 60 ocorrências, seguido pelo diálogo com 28, apresentação com 25 e, finalmente, a atividade de leitura foi a que registrou o menor número de erros, sendo apenas seis.

Figura 57 – Quantitativo dos erros sintáticos por atividade do *corpus* oral.



Fonte: a autora.

Na Figura 57, quanto o tipo de erro sintático, podemos observar que em todas as atividades onde os estudantes deviam criar uma produção própria e original, como é o caso das atividades de apresentação, debate e diálogo, o tipo de erro sintático apareceu numa quantidade bastante similar, sendo trinta ocorrências na atividade de debate, 26 na atividade de diálogo, e 22 na apresentação. À respeito à atividade de leitura, o número de ocorrências é consideravelmente menor, sendo apenas quatro o número de vezes que o erro foi encontrado, esse resultado pode dever-se a que não se trata de um tipo de atividade criativa e, portanto, as sentenças já estavam formuladas corretamente.

4.5.6 *Quantitativo dos “erros” por períodos no corpus escrito*

A seguir presentamos os dados correspondentes ao *corpus* escrito. Amostram-se na mesma ordem que a tabela anterior, ou seja, com três colunas que organizaram o período, o tipo de erro e o número de ocorrências encontradas.

Tabela 35 – Quantitativo das ocorrências por período e por tipo de erro do *corpus* escrito.

Período	Tipo de Erro	O°
2°	Léxico	9
2°	Morfológico	74
2°	Ortográfico	44
2°	Sintático	8
4°	Léxico	15
4°	Morfológico	35
4°	Ortográfico	20
4°	Sintático	8
6°	Léxico	8
6°	Morfológico	13
6°	Ortográfico	6
6°	Sintático	3
8°	Léxico	7
8°	Morfológico	11
8°	Ortográfico	7
8°	Sintático	3

Fonte: a autora.

Na Tabela 35, podemos apreciar os dados de ocorrências do *corpus* escrito, organizados por período. Cabe comentar que o *corpus* escrito contém um número inferior de palavras em comparação ao *corpus* oral, pelo que convém considerar cada um dos *corpus* como universos separados, quanto ao total de palavras e a diferença considerável, portanto, do número de erros.

A seguir apresentamos gráficos que apresentam os dados organizados por cada período.

Figura 58 – Quantitativo dos erros morfológicos por período do *corpus* escrito.



Fonte: a autora.

Na Figura 58, quanto ao tipo de erro morfológico, podemos apreciar que existe uma evolução no percurso do curso: no segundo período o número de ocorrências foi o mais alto, 74 vezes que se repetiu esse tipo de erro, seguido pelo quarto período, com 35 ocorrências, o sexto período com 13 e, finalmente, o menor número de ocorrências no oitavo período com apenas 11.

Figura 59 – Quantitativo dos erros ortográficos por período do *corpus* escrito.

Fonte: a autora.

Na Figura 59, quanto ao tipo de erro ortográfico, também é possível notar uma diminuição das ocorrências durante o percurso do Curso de Letras: no segundo período encontramos o número mais alto de ocorrências, chegando até 44, seguido pelo quarto período com vinte, sexto período com seis e oitavo período com sete, ou seja, mantendo-se nos dois últimos períodos estudados o número de erros do tipo ortográfico.

Figura 60 – Quantitativo dos erros lexicais por período do *corpus* escrito.



Na Figura 60, quanto ao tipo de erro léxico, podemos apreciar que a maior quantidade de ocorrências aconteceu no quarto período com quinze ocorrências, seguido do segundo período com nove, sexto período com oito e oitavo período com sete.

Figura 61 – Quantitativo dos erros sintáticos por período do *corpus* escrito.

Fonte: a autora.

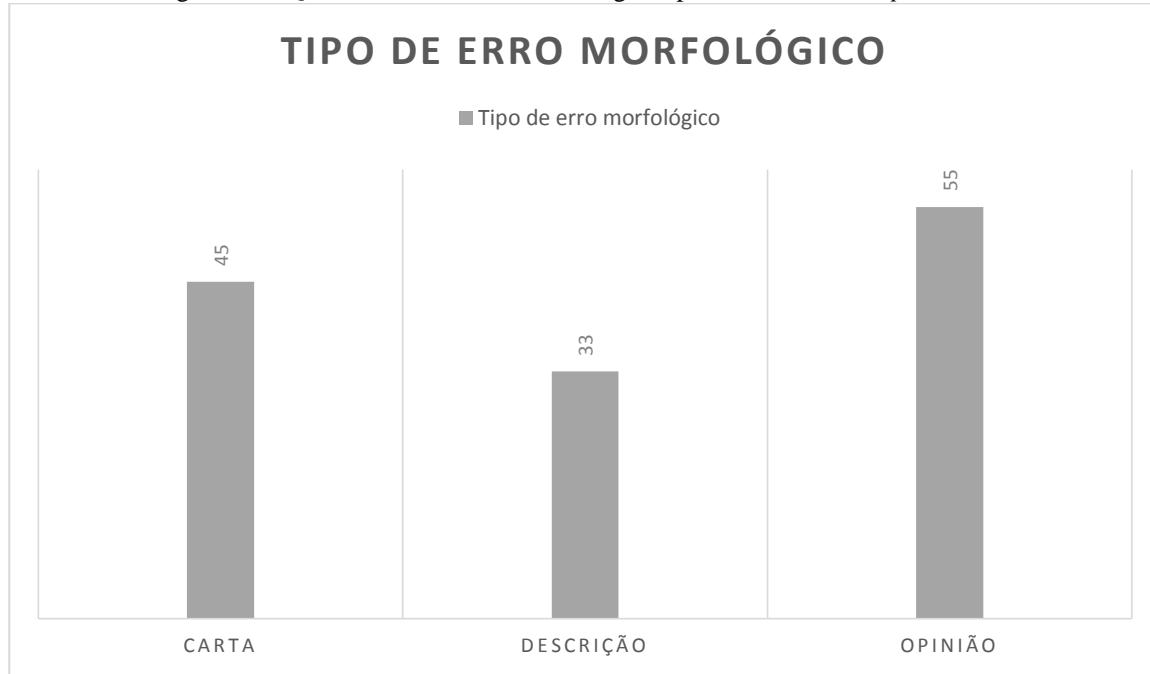
Na Figura 61, quanto ao tipo de erro sintático, podemos apreciar uma tendência à diminuir, pois nos primeiros dois períodos estudados, ou seja no segundo e no quarto período, o número de ocorrências foi de oito e, nos dois últimos períodos estudados, ou seja, sexto e oitavo período, o número de ocorrências foi apenas de três.

Tabela 36 – Quantitativo das ocorrências por atividade e por tipo de erro do *corpus*.

Atividade	Tipo de Erro	O°
Carta	Léxico	21
Carta	Morfológico	45
Carta	Ortográfico	28
Carta	Sintático	12
Descrição	Léxico	12
Descrição	Morfológico	33
Descrição	Ortográfico	13
Descrição	Sintático	7
Opinião	Léxico	6
Opinião	Morfológico	55
Opinião	Ortográfico	36
Opinião	Sintático	3

Fonte: a autora.

Na Tabela 36, podemos apreciar os dados de ocorrências de erros do *corpus* escrito, organizados por tipo de atividades. A seguir, apresentam-se gráficos que organizam a informação por tipo de erros.

Figura 62 – Quantitativo dos erros morfológicos por atividade do *corpus* escrito.

Fonte: a autora.

Na Figura 62, quanto ao tipo de erro morfológico, podemos apreciar que foi na atividade de opinião que se registrou o maior número de ocorrências, sendo elas um total de 55, seguido pela atividade “carta” com 45 ocorrências e, finalmente, a atividade de descrição com apenas 33 erros.

Figura 63 – Quantitativo dos erros ortográficos por atividade do *corpus* escrito.



Fonte: a autora.

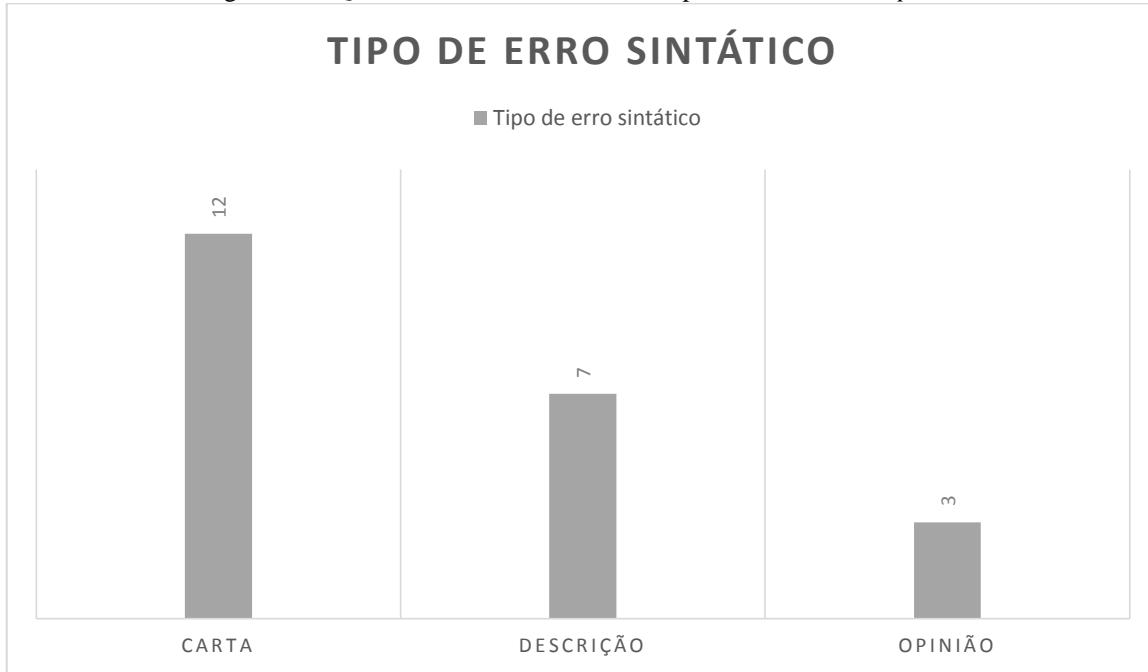
Na Figura 63, quanto ao tipo de erro ortográfico, podemos observar que a maior quantidade aconteceu na atividade de opinião com 36 ocorrências, seguido pela atividade de carta com 28 e, finalmente, a atividade de descrição com a menor quantidade de ocorrências desse tipo de erro, com apenas 13.

Figura 64 – Quantitativo dos erros lexicais por atividade do *corpus* escrito.

Fonte: a autora.

Na Figura 64, quanto ao tipo de erro léxico, podemos observar que o maior número de ocorrências está na atividade que precisava de mais capacidade criativa por parte do aluno, ou seja, na atividade de “carta”, com um número de 21 ocorrências. Na atividade, os alunos deviam basear-se numa situação hipotética na qual deviam escrever para alguém próximo desde um país estrangeiro, contando as impressões que o lugar provocou.

Figura 65 – Quantitativo dos erros sintáticos por atividade do *corpus* escrito.



Fonte: a autora.

Na Figura 65, quanto ao tipo de erro sintático, podemos apreciar que foi na atividade de “carta” que aconteceu a maior quantidade de ocorrências, totalizando doze, seguido pela atividade “descrição”, com sete ocorrências e, finalmente, a atividade com a menor quantidade de ocorrências foi a “opinião”, com apenas três registros.

4.6 Estudo qualitativo dos erros

Com relação ao estudo qualitativo dos erros, a seguir apresentamos a análise estruturada da seguinte maneira: primeiro separada por *corpus* oral e *corpus* escrito, assim, em cada um dos períodos encontraremos cinco exemplos que conterão o “erro”, a forma esperada para a palavra identificada como erro, e a possível causa.

4.6.1 Estudo qualitativo dos erros no *corpus* oral por períodos

Começando pelo segundo período estes foram os “erros” fonológicos produzidos pelos participantes e as possíveis causas. Como trata-se de um tipo de erro que teve 72 ocorrências, incluímos mais uma coluna para apresentar mais “erros” do mesmo tipo.

1. "...y sí el narcotráfico qui <FON [que]> incluso está agregado..."

Erro	Forma esperada	Possível causa	Outros erros do mesmo tipo
qui	que	Pronuncia da "e" final do espanhol com som da "i".	clubis (clubes), hispanohablanti (hispanohablante)

2. "... federáu <FON [federal]> de uberlányia..."

Erro	Forma esperada	Possível causa	Outros erros do mesmo tipo
federáu	federal	Pronuncia da "l" final (liquida, lateral) com som de "u".	legau (legal), confidencial (confidencial), espanhou (espanhol)

3. "... cinco miémbros que con másimo <FON [máximo]> de noventa..."

Erro	Forma esperada	Possível causa	Outros erros do mesmo tipo
másimo	máximo	Pronuncia da letra x com som de "s" (como em português), na posição inter vogal o som deviesse ser "ks".	testo (texto), prosimidad (proximidad), esistir (existir)

4. "... estoh <FON [estos]> son los principales..."

Erro	Forma esperada	Possível causa	Outros erros do mesmo tipo
estoh	estos	Aspiração da "s" final.	habilitadah (habilitadas), vierneh (viernes)

5. "... hehístren <FON [registren]> / no será obligatório..."

Erro	Forma esperada	Possível causa	Outros erros do mesmo tipo
hehístren	registren	Pronunciação da "r" aspirada.	hegistrado (registrado), heyistrarse (registrarse).

A seguir os erros do tipo lexical do segundo período do *corpus oral*

1. "... Continuará mésimo <LEX : SIN [de igual forma]>..."

Erro	Forma esperada	Possível causa
mesmo	de igual forma	Transferência do português.

2. "... puéde presentár canciones, filmes <LEX [películas]>..."

Erro	Forma esperada	Possível causa
filmes	películas	Transferência do português.

3. "...acá en el lanchonéte <LEX [kiosko]>..."

Erro	Forma esperada	Possível causa
lanchonete	kiosko	Transferência do português.

4. "...puéde presentár músicas <LEX [canciones]>..."

Erro	Forma esperada	Possível causa
músicas	canciones	Transferência do português.

5. "...no hay dúvida <LEX [duda]>..."

Erro	Forma esperada	Possível causa
duvida	duda	Transferência do português.

A seguir os erros do tipo morfológico do segundo período do *corpus oral*

1. "...tódas as <MOR [las]> escuélas..."

Erro	Forma esperada	Possível causa
as	las	Transferência do português.

2. "...inscribír pesonalménte <MOR [personalmente]>..."

Erro	Forma esperada	Possível causa
pesonalmemente	personalmente	Mistura a forma correta do português “pessoalmente” com a forma correta do espanhol “personalmente”.

3. "...se cómpra más <MOR [más]> marijuana..."

Erro	Forma esperada	Possível causa
mais	más	Transferência do português.

4. "...-películas- túdo <MOR [todo]>..."

Erro	Forma esperada	Possível causa
tudo	todo	Transferência do português.

5. "...lugáres son asín <MOR [así]>..."

Erro	Forma esperada	Possível causa
asín	así	Transferência do português.

A seguir os erros do tipo sintático do segundo período do *corpus oral*

1. "...me gústa múcho de <SIN [la]> música tambiéñ..."

Erro	Forma esperada	Possível causa
me gusta mucho de música	me gusta mucho la música	Em português: “gosto muito da música”. Existe uma transferência desde o português da preposição “de” e uma supressão do artigo.

2. "...prógráma de iniciación a <SIN : AD [a la]> docênciа..."

Erro	Forma esperada	Possível causa
iniciación a docencia	iniciación a la docencia	Em português: "programa de iniciação à docência". Existe transferência desde o português, pois "à" corresponde à conjunção de preposição "a" mais artigo "a", no caso do espanhol a forma correta de preposição mais artigo corresponde a: "a la".

3. "...principalménte al <SIN [en]> Brasíl..."

Erro	Forma esperada	Possível causa
principalmente al Brasil	principalmente en Brasil	Utilização incorreta da conjunção "al" (a + el). No caso o correto é o uso da preposição "en" que denota lugar.

4. "...acá na <SIN [en el]> lanchonéte..."

Erro	Forma esperada	Possível causa
acá na laanchonete	acá en el kiosko	Transferência do português da conjunção "na", em espanhol não existe tal conjunção, e a forma correta corresponde a "en el"

5. "...que nos comunicár <SIN [que comunicarnos]>..."

Erro	Forma esperada	Possível causa
que nos comunicar	que comunicarnos	Transferência do português, pois em espanhol a partícula do pronome pessoal vai depois do verbo.

A seguir os erros do tipo fonológico do quarto período do *corpus oral*

1. "...está autorizáo <FON [autorizado]>..."

Erro	Forma esperada	Possível causa
autorizao	autorizado	Supressão da "d" intervocálica.

2. "...quedará jegistrádo <FON [registrado]>..."

Erro	Forma esperada	Possível causa
jegistrado	registrado	Pronuncia da "r" do espanhol com som de "j" do espanhol. Possível transferência da pronuncia da "r" do português.

3. "...registro cugespondiente <FON [correspondiente]>..."

Erro	Forma esperada	Possível causa
cugespondiente	correspondiente	Pronuncia da "r" do espanhol com som de "g" do espanhol. Possível transferência da pronuncia do francês.

4. "...aquélas <FON [aquellas]> personas..."

Erro	Forma esperada	Possível causa
aquelas	aquellas	Pronuncia da "ll" com som de "l".

5. "...que la marijuána <FON [marihuana]>..."

Erro	Forma esperada	Possível causa
marijuana	marihuana	Pronuncia da sequência "hua" com som de "j" do espanhol.

A seguir os erros do tipo lexical do quarto período do *corpus oral*

1. "...yo soy una persóna cálma <LEX [tranquila]>..."

Erro	Forma esperada	Possível causa
calma	tranquila	Transferência do português.

2. "...a mi me gusta mûito <LEX [mucho]> ler..."

Erro	Forma esperada	Possível causa
muito	mucho	Transferência do português.

3. "...acá piérto <LEX [cerca]> del restoránte..."

Erro	Forma esperada	Possível causa
piérto	cerca	Transferência do português.

4. "...criánças <LEX [niños]> de dóce áños..."

Erro	Forma esperada	Possível causa
criánças	niños	Transferência do português.

5. "...a mi me gústa también asistír <LEX [ver]> películas..."

Erro	Forma esperada	Possível causa
asistir	ver	Transferência do português.

A seguir os erros do tipo morfológico do quarto período do *corpus oral*

1. "...éntre en vigéncia el mártes sés <MOR [seis]>..."

Erro	Forma esperada	Possível causa
ses	seis	Mistura da forma do francês "six" e do espanhol "seis".

2. "...mayóres de desócho <MOR [dieciocho]> años..."

Erro	Forma esperada	Possível causa
desócho	dieciocho	Mistura da forma do português “dezoito”, com a forma do espanhol “dieciocho”

3. "... el prodúto <MOR [producto]> no puéde estar..."

Erro	Forma esperada	Possível causa
produto	producto	Transferência do português.

4. "...determina la cuantidád <MOR [cantidad]>..."

Erro	Forma esperada	Possível causa
cantidad	cantidad	Mistura da forma do português “quantidade” e a forma no espanhol “cantidad”.

5. "...con algunos alúnos <MOR [alumnos]>..."

Erro	Forma esperada	Possível causa
alunos	alumnos	Transferência do português.

A seguir os erros do tipo sintático do quarto período do *corpus oral*

1. "... péro no me gustan náda de libros <SIN:AD [nada los libros]> de amor..."

Erro	Forma esperada	Possível causa
“no me gustan nada de libros”	“no me gustan nada los libros”	Mistura a sintaxe do português “nao gosto nada de libros” com a sintaxe do español “no me gustan nada los libros”.

2. "...pára viájes y tódo más <SIN [y todo lo demás]>..."

Erro	Forma esperada	Possível causa
“y todo más”	“y todo lo demás”	Transferência da expressão do português “e tudo mais”.

3. "...no téngo náda <SIN : AD [en contra]> contra..."

Erro	Forma esperada	Possível causa
"no tengo nada contra"	"no tengo nada em contra"	Transferência da expressão do português "não tenho nada contra".

4. "...sou <SIN [estoy]> a favór de medidas pára restringir..."

Erro	Forma esperada	Possível causa
"sou a favor"	"estoy a favor"	Transferência do português.

5. "...la situación financiera quédá difícil <SIN [se pone]>..."

Erro	Forma esperada	Possível causa
"queda difícil"	"se pone"	Transferência da sintaxe do português "fica difícil".

A seguir os erros do tipo fonológico do sexto período do *corpus oral*

1. "...que séa más asesible <FON [accesible]>..."

Erro	Forma esperada	Possível causa
asesible	accesible	No português brasileiro não tem palavras que possuam a combinação "cc" e a tendência é a pronunciar com som de "s"

2. "...el límite <FON [límite]>..."

Erro	Forma esperada	Possível causa
límite	límite	Transferência do português. Acentuação incorreta pro espanhol pois trata-se de uma palavra esdrúxula.

3. "...en el párque paseár con mi pero <FON [perro]>..."

Erro	Forma esperada	Possível causa
pero	perro	Pronuncia da "r" fraca em lugar da "rr" que é mais forte.

4. "...el primér país en jegulár <FON [regular]>..."

Erro	Forma esperada	Possível causa
jegular	regular	Pronúncia da "r" do espanhol com som de "j" do espanhol. Transferência do som da "r" em português.

5. "...recibirá un tíqueh <FON [tíquet]> con un número..."

Erro	Forma esperada	Possível causa
tíqueh	tíquet	Aspiração da "t" no final da sílaba.

A seguir os erros do tipo lexical do sexto período do *corpus oral*

1. "...no es la palabra ciérta <LEX [adecuada]>..."

Erro	Forma esperada	Possível causa
cierta	adecuada	Transferência do português.

2. "...úna un inflúyo <LEX [influencia]> pára que la mayoría de la población..."

Erro	Forma esperada	Possível causa
influyo	influencia	Inseguridade do falante, pois a forma em português corresponde a "influência" também, mas o falante acaba utilizando uma forma inventada.

3. "...entónces vá <LEX [anda]> preparándote..."

Erro	Forma esperada	Possível causa
vá	anda	Transferência do português.

4. "...las persoas que ultrapásen <LEX [sobrepasen]> ése límite..."

Erro	Forma esperada	Possível causa
ultrapasen	sobrepasen	Transferência do português.

5. "...sí y hay ún os múncho <LEX [muy]> problemáticos..."

Erro	Forma esperada	Possível causa
mucho	muy	Regras mal adquiridas. Antes de um adjetivo o correto é utilizar "muy".

A seguir os erros do tipo morfológico do sexto período do *corpus* oral

1. "...hásta cuatrocéntos < MOR [cuatrocéntos] ochénta grámos..."

Erro	Forma esperada	Possível causa
cuatrocéntos	cuatrocientos	Transferência do português.

2. "...hay ún a cóta < MOR [cuota]> ..."

Erro	Forma esperada	Possível causa
cota	cuota	Transferência do português.

3. "...péndo < MOR [pienso]> que la limitación..."

Erro	Forma esperada	Possível causa
penso	pienso	Transferência do português.

4. "...conóce los dádos < MOR [datos]> personáles..."

Erro	Forma esperada	Possível causa
dados	datos	Transferência do português.

5. "...éntre vénte < MOR [veinte]> y vendedós < MOR [veintidós]> pésos uruguáyos..."

Erro	Forma esperada	Possível causa
vente	veinte	Mistura das formas do português "vinte" e do espanhol "veinte".

A seguir os erros do tipo sintático do sexto período do *corpus oral*

No caso apresentaremos quatro exemplos, pois o número de erros desse tipo foi apenas cinco e um deles se repete.

1. "...le cuénte <SIN [cuéntele]> que los alúmnos..."

Erro	Forma esperada	Possível causa
le cuente	cuéntele	Transferência da sintaxe do português (lhe conte). A forma imperativa do espanhol coloca a partícula "le" no final do verbo e junto a ele.

2. "... y por vivír léjos extráño múcho a éllos <SIN [los extraño mucho]>..."

Erro	Forma esperada	Possível causa
extraño mucho a éllos	los extraño mucho	O pronome "ellos" deve ser substituído pela partícula "los"

3. "...por éso hácen con que <SIN : OM [hacen que]> los narcotraficantes..."

Erro	Forma esperada	Possível causa
hacen con que	hacen que	Transferência do português (fazem com que).

4. "...no me mánde alumnos tódos <SIN : OM [todos los]> instantes..."

Erro	Forma esperada	Possível causa
todos instantes	todos los instantes	Transferência do português (todo instante).

A seguir os erros do tipo fonológico do oitavo período do *corpus oral*

1. "...el contácto ya inmeyiáto <FON [inmediato]> con éllos..."

Erro	Forma esperada	Possível causa
inmeyiato	inmediato	Transferência da "d" do português.

2. "...pórque es una realidá <FON [realidad]>..."

Erro	Forma esperada	Possível causa
realidá	realidad	elisão da consoante final.

3. "... hay bários <FON [barrios]> entéros que tiénen escuélas..."

Erro	Forma esperada	Possível causa
bários	barrios	Pronuncia da "r" fraca do espanhol, no caso deviesse ter sido pronunciada a erre forte multivibrante.

4. "...del índice ihquierdo <FON [izquierdo]>..."

Erro	Forma esperada	Possível causa
ihquierdo	izquierdo	Aspiração do som de "s" na parte distensiva da sílaba.

5. "...la persóna inscrita debrá <FON [deberá]>..."

Erro	Forma esperada	Possível causa
debrá	deberá	Elisão da vogal "e".

A seguir os erros do tipo lexical do oitavo período do *corpus* oral

1. "...mismo <LEX [incluso]> aun con la prohibición del consumo..."

Erro	Forma esperada	Possível causa
mismo	incluso	Transferência do português (mesmo).

2. "...la marihuana es muy fláca <LEX [suave]>..."

Erro	Forma esperada	Possível causa
flaca	suave	Transferência da forma em português "fraca".

3. "...corrúptos van a tenér <LEX [haber]> en tódos los países..."

Erro	Forma esperada	Possível causa
tener	haber	Transferência do português. Em português a construção correta seria utilizando o verbo "ter".

4. "...luégo que encerré <LEX [finalicé]> la enseñánza média..."

Erro	Forma esperada	Possível causa
encerré	finalicé	Transferência do português.

5. "...por la mañána yo me acuérdo <LEX [despierto]> tempráno..."

Erro	Forma esperada	Possível causa
acuerdo	despierto	Transferência do português.

A seguir os erros do tipo morfológico do oitavo período do *corpus oral*

1. "...con fiscalización redobráda <MOR [redoblada]>..."

Erro	Forma esperada	Possível causa
redobráda	redoblada	Transferência do português.

2. "...dez <MOR [diez]> veinte reáles..."

Erro	Forma esperada	Possível causa
dez	diez	Transferência do português.

3. "...un momento <MOR [momento]> que va a querér más..."

Erro	Forma esperada	Possível causa
momento	momento	Forma inventada. Intuição do falante.

4. "...es úna necesidá <MOR [necesidad]>..."

Erro	Forma esperada	Possível causa
necesidá	necesidad	Supressão da "d" final.

5. "...persónas inocéntes morréndo <MOR [muriendo]> en las cálles..."

Erro	Forma esperada	Possível causa
morréndo	muriendo	Transferência do português.

A seguir os erros do tipo sintático do oitavo período do *corpus oral*

1. "...¿qué tal? ¿qué te opinas <SIN : OM [que opinas]> ?..."

Erro	Forma esperada	Possível causa
que te opinas	que opinas	Repetição redundante da partícula que dá conta da pessoa.

2. "...y puéde influenciár más <SIN: OM [a más]> persoñas a usar..."

Erro	Forma esperada	Possível causa
influenciar más personas	influenciar a más personas	Supressão da preposição "a".

3. "...sí cláro podémos <SIN : OM [nos podemos]> sentár un día..."

Erro	Forma esperada	Possível causa
podemos sentar	nos podemos sentar	Supressão do pronome "nos".

4. "...la guérra cóntre el narcotráfico se <SIN : OM [ha fallado]> ha fachado..."

Erro	Forma esperada	Possível causa
se há fallado	ha fallado	Inclusão desnecessária da partícula "se".

5. "...nádie es favorable <SIN : LEX [está a favor]> al consumo..."

Erro	Forma esperada	Possível causa
nadie es favorable	nadie está a favor	

4.6.2 Estudo qualitativo dos erros no *corpus* oral por períodos

A seguir os erros do tipo ortográfico do segundo período do *corpus* escrito

1. “...A mi me gusta más él <ORT [el]> español Argentino...”

Erro	Forma esperada	Possível causa
él	el	Confusão do acento “diacrítico”. “él” corresponde ao pronome pessoal, “el” ao artigo.

2. “por ser mas <ORT [más]> fácil...”

Erro	Forma esperada	Possível causa
mas	más	Ausência do acento “diacrítico”. “más” indica adição, “mas” indica oposição ou restrição.

3. “...A mi <ORT [mí]>me encantan...”

Erro	Forma esperada	Possível causa
mi	mí	Ausência do acento diacrítico. “mí” corresponde ao pronome pessoal, “mi” ao possesivo.

4. “...la assignatura <ORT [asignatura]> de español...”

Erro	Forma esperada	Possível causa
assignatura	asigtura	Mistura da forma do português “assinatura” com a do espanhol “asignatura”.

5. “...Los materiales tambien <ORT [también]> no son los mejores...”

Erro	Forma esperada	Possível causa
tambien	también	Ausência do acento, pois palavras “agudas” que acabam em “n”, “s” ou vogal, devem ser acentuadas graficamente.

A seguir os erros do tipo lexical do segundo período do *corpus* escrito

1. "...es bien desenvolvida <LEX [desarrollada]>..."

Erro	Forma esperada	Possível causa
desenvolvida	desarrollada	Transferência do português.

2. "...por que treino <LEX [entreno]> mucho..."

Erro	Forma esperada	Possível causa
treino	entreno	Transferência do português.

3. "...es una buena facultad <LEX [universidad]>..."

Erro	Forma esperada	Possível causa
facultad	universidad	Mistura da forma do português "faculdade" com a forma do espanhol "universidad".

4. "...en ensino <LEX [enseñanza]> de lenguas..."

Erro	Forma esperada	Possível causa
ensino	enseñanza	Transferência do português.

5. "...estoy acá lecionando <LEX[enseñando]>el Portugués..."

Erro	Forma esperada	Possível causa
lecionando	enseñando	Transferência do português.

A seguir os erros do tipo morfológico do segundo período do *corpus* escrito

1. "...já <MOR [ya]> que una lengua..."

Erro	Forma esperada	Possível causa
já	ya	Transferência do português.

2. "...que más escuto <MOR [escucho]>..."

Erro	Forma esperada	Possível causa
escuto	escucho	Transferência do português.

3. "...y cada falante <MOR [hablante]> prefiere usar..."

Erro	Forma esperada	Possível causa
falante	hablante	Transferência do português.

4. "...octubre <MOR [octubre]> 2009..."

Erro	Forma esperada	Possível causa
octubre	octubre	Mistura da forma do português "outubro" e a forma do espanhol "octubre".

5. "...me parecen adecuados <MOR [adecuados]>..."

Erro	Forma esperada	Possível causa
adecuados	adecuados	Transferência do português.

A seguir os erros do tipo sintático do segundo período do *corpus* escrito

1. "...siento falta <SIN [extraño]> de..."

Erro	Forma esperada	Possível causa
siento falta	extraño	Transferência da forma do português "sinto falta".

2. "...eso da mucho certo <SIN : MOR [resulta mucho]>..."

Erro	Forma esperada	Possível causa
da mucho certo	Resulta mucho	Transferência da forma do português "da muito certo"

3. "...estoy gustando <SIN [me está gustando]>..."

Erro	Forma esperada	Possível causa
estoy gustando	me está gustando	Transferência da forma do português "estou gostando"

4. "...actividades que mejor son <SIN [que son mejores]> para cada clase..."

Erro	Forma esperada	Possível causa
que mejor son	que son mejores	Comumente utiliza-se o verbo antes do adjetivo.

5. "...más complicado do que <SIN:AD[que]>en Brasil..."

Erro	Forma esperada	Possível causa
do que en Brasil	que en Brasil	Transferência do português da forma "do que".

A seguir os erros do tipo ortográfico do quarto período do *corpus* escrito

1. "...respecto del curso de ingles <ORT [inglés]>..."

Erro	Forma esperada	Possível causa
ingles	inglés	Falta o acento. "inglês" corresponde ao idioma, "inglês" à parte do corpo.

2. "...parte de los payses <ORT [países]>..."

Erro	Forma esperada	Possível causa
payses	países	Confusão no uso da "y" e "í".

3. "...tener seguridad linguistica <ORT [lingüística]>..."

Erro	Forma esperada	Possível causa
linguistica	lingüística	Se utiliza diéreses no espanhol nas palavras com combinação "gui" e "gue" quando o som da "u" deve ser pronunciado.

4. "...Ellos hablan con más fluides <ORT [fluidez]>..."

Erro	Forma esperada	Possível causa
fluides	fluidez	Confusão no uso d "s" e "z".

5.... "Sobre el português <ORT [portugués]> yo pienso..."

Erro	Forma esperada	Possível causa
português	portugués	Transferência do português, pois não existe o acento circunflexo em espanhol.

A seguir os erros do tipo lexical do quarto período do *corpus* escrito

1. "...Cada ciudad más formosa <LEX [hermosa]> que la otra..."

Erro	Forma esperada	Possível causa
formosa	hermosa	Transferência do português.

2. "...gran parte del pueblo es luchador y valoroso <LEX [valiente]>..."

Erro	Forma esperada	Possível causa
valoroso	valiente	Transferência do português.

3. "...Carísima <LEX [queridísima]> / Iracema..."

Erro	Forma esperada	Possível causa
carísima	queridísima	Transferência do português.

4. "...producir profesores y pesquisadores <LEX [investigadores]>..."

Erro	Forma esperada	Possível causa
pesquisadores	investigadores	Transferência do português.

5. "...Estoy con saudades <LEX [nostalgia]>..."

Erro	Forma esperada	Possível causa
saudades	nostalgia	Transferência do português.

A seguir os erros do tipo morfológico do quarto período do *corpus* escrito

1. "...Estados Unidos hay mucho <MOR [muchos]> más latinos..."

Erro	Forma esperada	Possível causa
mucho	muchos	Uso do singular de forma errada.

2. "...pois <MOR [pues]> pienso que la diversidade..."

Erro	Forma esperada	Possível causa
pois	pues	Transferência do português.

3. "...pero hay un sabor forte <MOR [fuerte]>..."

Erro	Forma esperada	Possível causa
forte	fuerte	Transferência do português.

4. "...existen muchas estelas <MOR [estrellas]>..."

Erro	Forma esperada	Possível causa
estrelas	estrellas	Transferência do português.

5. "...y contruyer <MOR [construir]> edificios verticales..."

Erro	Forma esperada	Possível causa
contruyer	construir	Palavra inventada, intuição do falante.

A seguir os erros do tipo sintático do quarto período do *corpus* escrito

1. "...Quien sabe yo permanezco más un tiempo <SIN [un tiempo más]> aqui..."

Erro	Forma esperada	Possível causa
más un tiempo aqui	un tiempo más aqui	Transferência da sintaxe do português.

2. "...estoy me <SIN [me estoy]> acostumbrando con la gente..."

Erro	Forma esperada	Possível causa
estoy me acostumbrando	me estoy acostumbrando	Transferência da sintaxe do português.

3. "...y está me gustando <SIN [me está gustando]> muchísimo mi empleo..."

Erro	Forma esperada	Possível causa
está me gustando	me está gustando	Confusão quanto a posição das partículas na sentença.

4. "...los profesores hacen con que <SIN : OM [hacen que]> los alumnos desarollen..."

Erro	Forma esperada	Possível causa
hacen con que	hacen que	Transferência da sintaxe do português.

5. "...son personas que les gustan <SIN : AD [gustan los]> brasileños..."

Erro	Forma esperada	Possível causa
les gustan brasileños	les gustan los brasileños	Omissão do artigo "los".

A seguir os erros do tipo ortográfico do sexto período do *corpus* escrito

Somente apresentaremos quatro exemplos para esse tipo de erro, pois alguns deles se repetem.

1. "...para darte informaciones de aquí <ORT [aquí]>..."

Erro	Forma esperada	Possível causa
aqui	aquí	Omissão do acento.

2. "...la particularidad de la lengua escojida <ORT [escogida]>..."

Erro	Forma esperada	Possível causa
escojida	escogida	Confusão no uso da "g" y "j" do espanhol.

3. "...solo por este pais <ORT [país]>..."

Erro	Forma esperada	Possível causa
país	país	Omissão do acento.

4. "...como parte de la economía <ORT [economía]>..."

Erro	Forma esperada	Possível causa
economia	economía	Omissão do acento.

A seguir os erros do tipo lexical do sexto período do *corpus* escrito

1. "...mismo <LEX [incluso]> echando de menos a todos..."

Erro	Forma esperada	Possível causa
mismo	incluso	Transferência do português.

2. "...Nada mejor para un aprendiente <LEX [aprendiz]>..."

Erro	Forma esperada	Possível causa
aprendiente	aprendiz	Transferência do português.

3. "...UFU soy <LEX [estoy]> encantada por este lugar..."

Erro	Forma esperada	Possível causa
soy	estoy	Transferência da forma do português “sou apaixonada por esse lugar”.

4. "...como estoy me saliendo <LEX : SIN [desempeñándome]>..."

Errome	Forma esperada	Possível causa
me saliendo	desempeñándome	A escolha do léxico foi inadequada, e a construção sintática feita por intuição.

5. "...15 alumnos por sala en <LEX [como]> máximo..."

Erro	Forma esperada	Possível causa
en	como	Utilização incorreta da preposição, pois “em” indica lugar.

A seguir os erros do tipo morfológico do sexto período do *corpus* escrito

1. "...programas gubiernamentales <MOR [gubernamentales]>..."

Erro	Forma esperada	Possível causa
gubiernamentales	gubernamentales	

2. "...me gusta el español messicano <MOR [mexicano]>..."

Erro	Forma esperada	Possível causa
messicano	mexicano	Troca “ss” por “x”, no português o som da “x” é de uma “s” do espanhol.

3. "...hay bueno <MOR [buenos]> profesores..."

Erro	Forma esperada	Possível causa
bueno	buenos	Construção correta pro plural é com “s” no final.

4. "...es una universidad de mucho respecto <MOR [respeto]>..."

Erro	Forma esperada	Possível causa
respecto	respeto	Transferência do português.

5. "...y de darles herramientas <MOR [herramientas]> que posibiliten..."

Erro	Forma esperada	Possível causa
herramientas	herramientas	Inclusão desnecessária da letra "í" em "hierra"

A seguir os erros do tipo sintático do sexto período do *corpus* escrito

Somente apresentaremos três exemplos, pois foram as únicas ocorrências.

1. "...multimedia para <SIN : AD [para que]> los profesores den sus clases..."

Erro	Forma esperada	Possível causa
para los	para que los	Omissão do da conjunção "que".

2. "...vine estudiar en <SIN : AD [en la]> UFU..."

Erro	Forma esperada	Possível causa
en UFU	en la UFU	Omissão do artigo.

3. "...creo que a la causa <SIN : OM [a causa]> de tanta tarea..."

Erro	Forma esperada	Possível causa
a la causa	a causa	Utilização desnecessária do artigo "la".

A seguir os erros do tipo ortográfico do oitavo período do *corpus* escrito

1. "...Tiene una buenisima <ORT [buenísima]> biblioteca..."

Erro	Forma esperada	Possível causa
buenisima	buenísima	Ausência do acento esdrúxulo.

2. "...y está dibidido <ORT [dividido]> en cuatro..."

Erro	Forma esperada	Possível causa
dibidido	dividido	Utilização incorreta da "b", talvez a causa da pouca diferenciação que se faz no espanhol na pronuncia da "b" e "v".

3. "...Hola mamá / que <ORT [qué]> onda..."

Erro	Forma esperada	Possível causa
que onda	qué onda	Ausência do acento, pois é uma pergunta exclamativa.

4. "...principalmente es la ultima <ORT [última]>..."

Erro	Forma esperada	Possível causa
ultima	última	Ausência do acento esdrúxulo.

5. "...porque la unica <ORT [única]> manera de tener acceso..."

Erro	Forma esperada	Possível causa
unica	única	Ausência do acento esdrúxulo.

A seguir os erros do tipo lexical do oitavo período do *corpus* escrito

1. "...La lengua es daptada <LEX [dotada]> de una riqueza inmensurable..."

Erro	Forma esperada	Possível causa
daptada	dotada	Intuição por parte do falante.

2. "...el hace su escoja <LEX [elección]>...."

Erro	Forma esperada	Possível causa
escoja	elección	Transferência da forma do português "fazer sua escolha"

3. "...el alumno hay <LEX [tiene]> que hacer obligatoriamente..."

Erro	Forma esperada	Possível causa
hay	tiene	Transferência do português, "o aluno tem que fazer".

4. "...dentro de dos años me vuelvo para <LEX [con]> ustedes..."

Erro	Forma esperada	Possível causa
para	con	"Para" indica propósito ou finalidade, "con" indica relação de associação.

5. "...no tenemos asignaturas de lengua lo bastante <LEX [suficiente]>..."

Erro	Forma esperada	Possível causa
bastante	suficiente	O uso da palavra bastante é semanticamente lógico, mas o uso comum seria da palavra "suficiente" no caso.

A seguir os erros do tipo morfológico do oitavo período do *corpus* escrito

1. "...poder ver sus cuadros de acerca <MOR [cerca]> ..."

Erro	Forma esperada	Possível causa
acerca	cerca	Confusão na construção da palavra, pois "acerca" significa "sobre" e "cerca", "perto".

2. "...en nuestra casa y <MOR [e]> informales que dentro de dos años..."

Erro	Forma esperada	Possível causa
y	e	No caso, a utilização do conector correta deve ser um "e", pois a palavra que continua depois do conector começa com letra "i".

3. "...Hoy empecé em <MOR [en]> mi nuevo trabajo..."

Erro	Forma esperada	Possível causa
em	en	Transferência do português.

4. "...es una conduta <MOR [conducta]> irregular o inadecuada..."

Erro	Forma esperada	Possível causa
conduta	conducta	Transferência do português.

5. "...un organismo que defenda <MOR [defienda]> politicamente..."

Erro	Forma esperada	Possível causa
defenda	defienda	Transferência do português.

A seguir os erros do tipo sintático do oitavo período do *corpus* escrito

Nessa parte somente se apresentarão dois exemplos:

1. "...que no son iguales los de <SIN : AD [a los de]> europa..."

Erro	Forma esperada	Possível causa
iguales los de europa	iguales a los de europa	Ausência da preposição "a".

2. "A mí me gustan mucho de estas <SIN : OM [mucho estas]>..."

Erro	Forma esperada	Possível causa
me gustan mucho de estas	me gustan mucho estas	Transferência do português, "gosto muito dessas".

Os "erros" que tomamos como referência para exemplificar a pressente sessão apresentaram-se separados por tipo de *corpus*, numerados por período e por tipo de "erro", incluindo a forma esperada da palavra e a possível causa de cada "erro".

A possível causa surge com o intuito de apresentar uma hipótese do que poderia provocar o "erro", sem pretender ser a resposta exata da ocorrência de cada erro, pois isso corresponderia ao resultado dum trabalho a ser desenvolvido caso por caso com cada um dos participantes da pesquisa, e que por tanto, levaria por diante um trabalho minucioso, mais personalizado e intensivo. Para a presente sessão decidimos abranger somente o nível duma causa hipotética, constituindo-se como uma sessão de caráter analítico descriptivo.

5 CONSIDERAÇÕES FINAIS

A motivação dessa pesquisa, como referido no início desse trabalho, começou como uma motivação pessoal como professora de castelhano e, também, pela minha condição de estudante estrangeira nesse país luso falante. Uma das minhas dúvidas iniciais foram precisamente, qual seria o interesse por parte dos aprendizes brasileiros pela língua espanhola e, além disso, onde eles teriam o contato inicial com esta. No momento prévio à conformação do grupo de participantes que iriam realizar as atividades, pudemos constatar, por meio do questionário de sondagem, onde os aprendizes teriam o seu primeiro contato com a língua espanhola. As respostas foram variadas, mas a maior parte delas indicaram que o aprendizado da língua ocorreu principalmente durante o percurso do Curso de Letras, utilizando livros, músicas e filmes como material auto didático e também pelo contato com estrangeiros.

A respeito da extensão do *corpus* oral é importante mencionar que o tamanho do *corpus* oral foi maior, com 15.725 itens, enquanto o *corpus* escrito teve apenas 5.173.

Ao considerar a extensão do *corpus* oral por períodos, é pertinente mencionar que a produção foi incrementando de maneira gradual desde o segundo até o oitavo período. Isto se deve possivelmente ao número de horas da especialidade de espanhol as quais os participantes foram expostos, o que fez com que os alunos se mostrassem cada vez mais confiantes quanto a produção à medida que eles avançassem de período no curso de Letras, deste modo, a produção foi mais prolífica ao atingir o último período do curso de Letras.

No caso oposto da extensão do *corpus* oral, se apresenta o caso da atividade de “apresentação”, a qual mostrou o número mais baixo de itens, sendo apenas 2.111, resultado que possivelmente foi causa da dificuldade por parte dos participantes de desenvolverem o tópico proposto, o qual consistia principalmente em falarem de si próprios.

Quanto as palavras mais recorrentes no *corpus* oral e escrito, cabe mencionar que, o intuito principal foi oferecer um panorama geral do *corpus* de estudo que foi trabalhado, nesse sentido, podemos apreciar que, a lista de palavras lexicais que resultaram ser as mais recorrentes, não são exclusivamente erros, ela representa as palavras lexicais mais utilizadas em geral, convertendo-se em uma amostra geral do repertório dos participantes.

Quanto a extensão do *corpus* oral por atividade, é pertinente fazer um destaque na atividade de leitura, a qual se mostra como a atividade com o maior número de itens, aportando 7.110. É importante esclarecer que a atividade de leitura não possui um caráter criativo, pois representa a reprodução oral de uma produção já existente, a qual os participantes simplesmente

reproduziram oralmente (leram). Por tanto a atividade com maior produção de itens foi o “debate”, aportando 3.773, atividade na qual os participantes se mostraram mais confiantes e dispostos em formularem opiniões e argumentações.

Nesse momento final da redação da pesquisa, também é possível desvendar as perguntas que nortearam o presente estudo. A primeira delas fazia referência a quais eram os “erros” mais frequentes na fala e na escrita de aprendizes de ELE. Quanto ao *corpus* oral, os erros mais frequentes foram os morfológicos, com um total de 546 erros, seguidos pelos erros de tipo fonológicos com uma total de 316. Quanto ao *corpus* escrito, os erros mais recorrentes também foram os do tipo morfológico com um total de 133 ocorrências, seguidos pelos erros do tipo ortográfico com um total de 77.

Quanto ao porquê das ocorrências dos erros, entre as possíveis causas, podemos dizer que, na maior parte dos casos, tratou-se de erros por interferência da LM dos aprendizes. Eles apareceram em todos os níveis de análise propostos e etiquetados, ou seja, nos erros do tipo morfológico, fonológico, ortográfico, lexical, e sintático.

Quanto a sessão de análise qualitativo, foi constituída como um análise descritivo, fazendo um destaque na inclusão duma possível causa de cada erro exemplificado, sem ter o intuito de ser a resposta exata para cada ocorrência do erro, mas apresentando-se como uma hipótese da causa da ocorrência.

A utilização da LC como abordagem para o desenvolvimento dessa pesquisa e, mais especificamente, a utilização do programa *WordSmith Tools*, foram essenciais para poder tratar do assunto da interlíngua, pois, com essa ferramenta, foi possível processar os documentos que continham as produções das atividades, assim como obter dados, permitindo organizá-los e arquivá-los de diversas formas, outorgando liberdade à pesquisadora, quanto a ordem e ao tratamento que, desses dados, se pudessem realizar.

As análises desenvolvidas no presente trabalho constituem-se como uma contribuição em potencial, pois, a partir dos resultados do mesmo, será possível por exemplo, tomar em conta quais os erros mais frequentes, e tentar implementar ações pedagógicas as quais incentivem um novo plano de ensino, que tenha como intuito de melhorar as áreas que apresentaram mais debilidades.

Outro uso potencial desse estudo, e tomando em conta, as múltiplas possibilidades a serem abordadas no *corpus* compilado, seria servir como base para outras futuras pesquisas.

É fundamental mencionar quais foram as limitações principais do presente trabalho. A primeira delas, foi o fato de que como aluna estrangeira, eu pudesse atingir um nível de língua

portuguesa avançado que permitisse a redação do presente trabalho.

Outra das limitações, e a que pessoalmente considero como mais importante, foi o fator tempo, pois até o segundo semestre cursado não tive orientador, sendo no período de um ano e meio que realizamos múltiplas ações: redatamos o projeto de pesquisa, compilamos o *corpus* de estudo (planificação e aplicação das atividades, digitação e transcrição de arquivos, revisão dos arquivos), etiquetamos o *corpus*, leitura de referencial teórico e escrita do texto de dissertação, enquanto ainda era preciso cursar disciplinas para cumprir com os créditos que o PPGEL exige. Sinceramente considero que foi um desafio ao tempo e as condições estabelecidas, que afortunadamente pudemos concretizar cumprindo com todos os prazos estipulados.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRADE, O. *Interlíngua oral e léxico de brasileiros aprendizes de espanhol*. Londrina: Eduel, 2011.
- BARALO, M. *La adquisición del español como lengua extranjera*. Madrid: Arco Libros, 1999.
- BARALO, M. La interlengua del hablante no nativo. In: SÁNCHEZ LOBATO, J.; SANTOS GARGALLO, I. (Coord.). *Vademécum para la formación de profesores de español como segunda lengua (L2) / lengua extranjera (LE)*. Madrid: SGEL, 2004. p. 369-389.
- BORBA, F. S. *Dicionário de usos do português do Brasil*. São Paulo: Ed. Ática, 2002.
- BOSQUE, I.; DEMONTE, V. (eds.). *Gramática descriptiva de la lengua española*. 3 vols. Madrid: Real Academia Española / Espasa Calpe, 1999.
- BERBER SARDINHA, T. *Lingüística de Corpus*. Barueri, SP: Manole, 2004.
- BORDÓN MARTÍNEZ, T. La evaluación de la expresión oral y de la comprensión auditiva. In: SÁNCHEZ LOBATO, Jesús; SANTOS GARGALLO, Isabel (Coord.). *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2) / lengua extranjera (LE)*. Madrid: SGEL, 2004. p. 983 - 1003.
- CAMARENA, E. Análisis de Interlengua en un *Corpus* de portugués escrito de estudiantes castellano hablantes universitarios. *Revista de Lingüística y Lenguas Aplicada*, N°6. Universidad Politécnica de Valencia: Departamento de Lingüística Aplicada: Valencia, 2011, p. 71-82.
- CRUZ PIÑOL, M. *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros, 2012.
- DURÃO, A. *Ánalisis de Errores en la interlengua de brasileños aprendices de español y de españoles aprendices de portugués*. 2^a ed. Londrina: Eduel, 2004.
- DURÃO, A. *La interlengua*. Madrid: Arco Libros, 2007.
- FERNÁNDEZ PÉREZ, M. *Lingüística de corpus y adquisición de la lengua*. Madrid: Arco Libros, 2011.
- GALLARDO PAÚLS, B. *Comentario de textos conversacionales: I. De la teoría al comentario*. Madrid: Arco Libros, 1998.
- GIL-TORESANO BERGES, M. La comprensión auditiva. In: SÁNCHEZ LOBATO, Jesús; SANTOS GARGALLO, Isabel (Coord.). *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2) / lengua extranjera (LE)*. Madrid: SGEL, 2004. p. 899 - 915.
- GIOVANNINI, A; et al. *Profesor en acción*. Madrid: Edelsa, 1996.

GLYNN, D. *et al*, *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Göttingen: Hubert & Co. GmbH & Co. KG, 2010.

HALLIDAY, M. A. K. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edgard Arnold, 1978.

KENEDY, E. *Curso básico de linguística gerativa*. São Paulo: Contexto, 2013.

LICERAS, J. M. *La adquisición de las lenguas extranjeras: hacia un modelo de análisis de la interlengua*. Madrid: Visor, 1992.

LLANOS, L. *Corpus Oral de Español como Lengua Extranjera: Guía didáctica*. Laboratorio de Lingüística Informática: Universidad Autónoma de Madrid, 2013.

MASIP, V. *Gramática histórica portuguesa e espanhola: um estudo sintético e contrastivo*. São Paulo: EPU, 2003.

MOLINER, M. *Diccionario de uso del español*. Ed. Electrónica 3.0. Barcelona: Gredos 2009.

NEVES, M. H. De M. *Gramática de usos do português*. São Paulo: UNESP, 2000.

NOVODVORSKI, A.; HERRERA, A. M. F. *Conversaciones con un lingüista de corpus*: Profesor Dr. Giovanni Parodi. *Letras & Letras*, Vol. 30, Nº 2. Edufu: Uberlândia, 2014, p. 452-466.

PARODI, G. *Lingüística de Corpus: de la teoría a la empiria*. Madrid: Iberoamericana / Vervuert, 2010.

PERINI, M. A. *Principios de lingüística descriptiva: introdução ao pensamento gramatical*. São Paulo: Parábola Editorial, 2006.

PINILLA GÓMEZ, R. La expresión oral. In: SÁNCHEZ LOBATO, J.; SANTOS GARGALLO, I. *Vademécum para la formación de profesores: enseñar español como segunda lengua (L2) / lengua extranjera (LE)*. Madrid: SGEL, 2004. P. 879 – 897.

QUILIS, A. *Principios de fonología y fonética españolas*. 8^a ed. Madrid: Arco Libros, 2008.

REAL ACADEMIA ESPAÑOLA; ASOCIACIÓN DE ACADEMIAS DE LA LENGUA ESPAÑOLA. *Nueva gramática de la lengua española: Fonética y Fonología*. Barcelona: Espasa Libros, 2011.

RUFAT, A. *El verbo dar en español escrito de aprendices de L1 inglés: estudio comparativo entre hablantes no nativos y hablantes nativos basado en corpus*. Universidad de Extremadura.

SANTOS GARGALLO, I. El análisis de errores en la interlengua del hablante no nativo. In: SÁNCHEZ LOBATO, J.; SANTOS GARGALLO, I. (Coord.). *Vademécum para la formación de profesores de español como segunda lengua (L2) / lengua extranjera (LE)*. Madrid: SGEL, 2004. P. 391-410.

- SILVA, J. *El léxico en las producciones orales de aprendices brasileños de español bajo la óptica de la Lingüística de Corpus*.
- SÖHRMAN, I. *La lingüística contrastiva como herramienta para la enseñanza de lenguas*. Madrid: Arco Libros, 2007.
- VAQUERO, M. de Ramírez. *El español de América I: Pronunciación*. 3^a edição. Madrid: Arco Libros, 2003.
- VÁSQUEZ, G. *¿Errores? ¡Sin Falta!* Madrid: Edelsa Grupo Didascalia, S.A, 1998.
- VÁSQUEZ, G. El concepto de error: estado de la cuestión y posibles investigaciones. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*. N° 5, Año 2009. Universidad de Nebrija, 2009a. p. 104-112.
- VÁSQUEZ, G. Análisis de errores, el concepto de corrección y el desarrollo de la autonomía. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*. N° 5, Año 2009. Universidad de Nebrija, 2009b. p. 113-122.
- TAGNIN, S. E. O. Glossário em Linguística de *Corpus*. In: VIANA, V; TAGNIN, S. E. O. *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2011. p. 357-361.
- YOKOTA, R.A *marcação de caso acusativo na interlíngua de brasileiros que estudam espanhol*. Dissertação (Mestrado em Estudos Linguísticos). São Paulo: FFLCH/USP, 2001

APÊNDICE A: Questionário de Sondagem aplicado aos estudantes do Curso de Letras/Espanhol.

QUESTIONÁRIO DE SONDAGEM¹

Nome completo:

Idade: Sexo: M () H () Língua materna:

Período letivo no momento da pesquisa: 2º () 4º () 6º () 8º ()

Responda as seguintes perguntas assinalando segundo estes critérios:

1 = Nulo; 2 = Pouco; 3 = Suficiente; 4 = Proficiente

1.1. Que nível de conhecimento considera que tinha quando você iniciou o Curso de Letras?

	Audição	Leitura	Escrita	Fala
Língua Espanhola:	()	()	()	()

1.2. Que nível de conhecimento considera ter atualmente?

	Audição	Leitura	Escrita	Fala
Língua Espanhola:	()	()	()	()

2. Marque a(s) opção(ões) que mais representa(m) seus interesses:

2.1. De que maneira você aprendeu espanhol?

- () Na escola: ensino médio ou fundamental
- () Curso de idiomas
- () Por meio de livros, música e/ou filmes
- () Contato com estrangeiros
- () Viagens
- () Outro – Especifique

2.2. Quais são as principais dificuldades que você encontra em espanhol?

- () Pronúncia
- () Gramática
- () Vocabulário
- () Compreensão
- () As semelhanças entre as duas línguas me confunde

2.3. Minha motivação pelo estudo da língua espanhola reside em que

- () é a língua falada nos países vizinhos do Brasil.
- () é um importante recurso para o mercado de trabalho.
- () é uma língua em ascensão no mundo.

2.4. Que utilidades práticas vê na aprendizagem de espanhol?

- () É uma língua útil para o turismo e para a comunicação internacional
- () É uma língua instrumental: dá acesso a bibliografia, a fontes de pesquisa, jornais, filmes, etc.
- () É uma língua que abre novos horizontes no mercado de trabalho

¹Questionário adaptado de Durão (2004, p. 321).

APÊNDICE B: Termo de Consentimento Livre e Esclarecido assinado pelos participantes da pesquisa.

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(maiores de 18 anos)

Você está sendo convidado/a para participar da pesquisa intitulada **ESTUDO CONTRASTIVO DA INTERLÍNGUA EM CORPUS ORAL E ESCRITO DE APRENDIZES DE ELE**, sob a responsabilidade do professor pesquisador Ariel Novodvorski e da estudante de Mestrado em Estudos Linguísticos Ana Maria Fritz Herrera. Nesta pesquisa, procura-se analisar aspectos da interlíngua de estudantes de espanhol falantes de português, num corpus de produções orais e escritas, como um meio de identificar as principais dificuldades na aprendizagem dessa língua estrangeira no Brasil.

Em caso de querer fazer parte da pesquisa e de ser selecionado/a, segundo as informações apresentadas no *Questionário de sondagem*, você participará da gravação de quatro produções orais e também da produção de três textos escritos, atividades que serão realizadas no Laboratório de Línguas (LABLING) do Instituto de Letras e Linguística da Universidade Federal de Uberlândia. As atividades realizadas na instituição de ensino serão conduzidas pela estudante de mestrado Ana Maria Fritz Herrera. As produções orais consistirão em uma apresentação, um debate, um diálogo e uma leitura. As produções escritas serão uma carta, um texto opinativo e uma descrição. Todas as produções ficarão sob a responsabilidade da estudante de mestrado Ana Maria Fritz Herrera, sob a supervisão do seu orientador, professor Ariel Novodvorski, que se comprometem a guardar as informações que serão analisadas. As informações apresentadas no *Questionário de sondagem* serão utilizadas para seleção dos participantes e, posteriormente, para fins exclusivos de pesquisa, guardando o devido sigilo, no concerente às informações de cunho pessoal. Em nenhum momento você será identificado/a. Tanto o professor pesquisador como sua assistente se comprometem a manter sigilo absoluto.

O áudio das gravações será transscrito e as produções escritas digitalizadas e digitadas. Após a transcrição dos dados, as gravações serão apagadas e as produções escritas devolvidas aos participantes da pesquisa. Os resultados da pesquisa serão publicados e, ainda assim, a identidade dos participantes será preservada. Você não terá nenhum gasto e/ou ganho financeiro por participar da pesquisa. O único risco existente é o de você se sentir constrangido/a em expor algum assunto de modo mais detalhado nas produções orais e escritas. Os benefícios serão indiretos, uma vez que, a partir da análise dos dados, você poderá conhecer melhor as dificuldades mais recorrentes na aprendizagem da língua espanhola no Brasil.

Você é livre para deixar de participar desta pesquisa a qualquer momento sem nenhum prejuízo para você ou mesmo para a pesquisa feita. Uma cópia deste Termo de Consentimento Livre e Esclarecido ficará com você. Qualquer dúvida a respeito da pesquisa, você poderá entrar em contato com o Pesquisador responsável: Prof. Dr. Ariel Novodvorski (Professor efetivo do ILEEL/UFU). Av. João Naves de Ávila, nº 2121, Bloco 1G, sala 249, Campus Santa Mônica – Uberlândia – MG. CEP: 38408-100. Fone: 34-3239-4314. Poderá também entrar em contato com o Comitê de Ética na Pesquisa com Seres-Humanos – Universidade Federal de Uberlândia: Av. João Naves de Ávila, nº 2121, bloco A, sala 224, Campus Santa Mônica – Uberlândia – MG, CEP: 38408-100; fone: 34-3239-4131.

Eu aceito participar do projeto citado acima, voluntariamente, após ter sido devidamente esclarecido.

Participante da pesquisa

Uberlândia,de..... de 2014

Assinatura do pesquisador

Prof. Dr. Ariel Novodvorski

ANEXO A: Parecer Consustanciado do CEP



UNIVERSIDADE FEDERAL DE
UBERLÂNDIA/MG



Continuação do Parecer: 935.682

termos de extensão das amostras e da diversidade em número de expoentes por período. Os pesquisadores esperam poder contribuir no aprimoramento do ensino e aprendizagem da língua espanhola como língua estrangeira para estudantes brasileiros, por meio de um inventário dos erros mais frequentes, das tipologias e suas prováveis causas.

METODOLOGIA: Esta pesquisa se propõe analisar um corpus oral e escrito de estudantes universitários de espanhol, falantes de português brasileiro, no Instituto de Letras e Linguística da Universidade Federal de Uberlândia, Minas Gerais. Participarão da pesquisa apenas quatro (4) estudantes por período (2º, 4º, 6º e 8º períodos letivos), totalizando dezessete (17) participantes. Os primeiros instrumentos serão os desenhados para coletar dados orais. Para este propósito, será utilizado o Laboratório de Línguas (LABLING), onde serão gravadas produções como leituras, apresentações, debates e diálogos. Serão aplicadas as mesmas atividades a todos os participantes, em todos os períodos. Desse modo, cada estudante fará uma leitura, uma apresentação, participará de um diálogo com outro colega (feito em dupla com colega de seu grupo) e ainda participará de um debate (feito no grupo relativo a seu período). Ao término das produções orais, haverá 44 produções orais. Para a parte escrita da coleta de dados, serão realizadas três atividades, em datas e local combinado previamente com os estudantes. Os participantes produzirão à mão, em papel e a caneta, uma carta, um texto de caráter opinativo e outro descritivo. Não haverá consulta a dicionários. As atividades serão realizadas em três dias diferentes, um para cada uma das atividades diferentes de escrita. Ao término das produções escritas, haverá um total de 48 textos. As gravações das produções orais e escritas serão realizadas no Laboratório de Línguas (LABLING) do Instituto de Letras e Linguística da Universidade Federal de Uberlândia. Serão feitas suas correspondentes transcrições ortográficas e digitalização dos textos escritos, com o objetivo de analisar e descrever a tipologia de erros mais frequentes dos estudantes de espanhol. Para as gravações, serão utilizados os recursos e programas já existentes no LABLING. A equipe pesquisadora visitará as turmas do Curso de Graduação em Letras/Espanhol, fará uma breve apresentação do projeto e entregará aos alunos interessados uma cópia do Questionário que será aplicado aos estudantes. Posteriormente, o grupo de informantes selecionado receberá uma cópia do TCLE, para coleta de assinatura, após um período de leitura e eventuais esclarecimentos. O local definido para a devolução do TCLE assinado é o Bloco 5M sala 3 do Campus Santa Mônica, nas quintas-feiras entre 14h e 16h, momento em que ocorrerão as reuniões com o orientador do presente projeto. Os participantes responderão um questionário com informações básicas relativas à idade, gênero, procedência, tempo de estudo da língua espanhola antes do ingresso ao Curso superior, em caso de se aplicar.

E

d

L

Endereço: Av. João Naves de Ávila 2121- Bloco "1A", sala 224 - Campus Sta. Mônica

Bairro: Santa Mônica

CEP: 38.408-144

UF: MG

Município: UBERLÂNDIA

Telefone: (34)3239-4131

Fax: (34)3239-4335

E-mail: cep@propp.ufu.br

Página 02 de 06

Endereço: Av. João Naves de Ávila 2121- Bloco "1A", sala 224 - Campus Sta. Mônica

Bairro: Santa Mônica

CEP: 38.408-144

UF: MG

Município: UBERLÂNDIA

Telefone: (34)3239-4131

Fax: (34)3239-4335

E-mail: cep@propp.ufu.br

Página 01 de 06



Continuação do Parecer: 235.582

BENEFÍCIOS: A pesquisa pretende trazer uma contribuição para o ensino e aprendizagem de espanhol no Brasil, a partir do levantamento dos erros mais frequentes, num corpus oral e escrito de espanhol como língua estrangeira. Por outro lado, também se procura avançar nas pesquisas que envolvem a análise de corpus oral e escrito de aprendizes de língua estrangeira. Os resultados desta pesquisa poderão ser de utilidade para a elaboração de materiais didáticos, pensados a partir das principais dificuldades encontradas nas produções dos estudantes de espanhol no Curso de Letras, futuros professores da língua no Brasil.

Comentários e Considerações sobre a Pesquisa:

A pesquisa em questão demonstra relevância acadêmica.

Considerações sobre os Termos de apresentação obrigatória:

Termos devidamente anexados. Cronograma adequado.

Recomendações:

Não há.

Conclusões ou Pendências e Lista de Inadequações:

As pendências apontadas no parecer 922.583 foram atendidas.

De acordo com as atribuições definidas na Resolução CNS 466/12, o CEP manifesta-se pela aprovação do protocolo de pesquisa proposto.

O protocolo não apresenta problemas de ética nas condutas de pesquisa com seres humanos, nos limites da redação e da metodologia apresentadas.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Data para entrega de Relatório Final ao CEP/UFU: Fevereiro de 2016.

OBS.: O CEP/UFU LEMBRA QUE QUALQUER MUDANÇA NO PROTOCOLO DEVE SER INFORMADA IMEDIATAMENTE AO CEP PARA FINS DE ANÁLISE E APROVAÇÃO DA MESMA.

Endereço: Av. João Naves de Ávila 2121 - Bloco "1A", sala 224 - Campus Sta. Mônica
 Bairro: Santa Mônica CEP: 38.400-144
 UF: MG Município: UBERLÂNDIA
 Telefone: (34)3239-4131 Fax: (34)3239-4335 E-mail: cep@propp.ufu.br



Continuação do Parecer: 995.982

O CEP/UFU lembra que:

- a- segundo a Resolução 466/12, o pesquisador deverá arquivar por 5 anos o relatório da pesquisa e os Termos de Consentimento Livre e Esclarecido, assinados pelo sujeito de pesquisa.
- b- poderá, por escolha aleatória, visitar o pesquisador para conferência do relatório e documentação pertinente ao projeto.
- c- a aprovação do protocolo de pesquisa pelo CEP/UFU dá-se em decorrência do atendimento a Resolução CNS 466/12, não implicando na qualidade científica do mesmo.

Orientações ao pesquisador:

- O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Res. CNS 466/12) e deve receber uma via original do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado.
- O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Res. CNS 466/12), aguardando seu parecer, exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade de regime oferecido a um dos grupos da pesquisa que requeram ação imediata.
- O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Res. CNS 466/12). É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.
- Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. Em caso de projetos do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma, junto com o parecer aprobatório do CEP, para serem juntadas ao protocolo inicial (Res. 251/97, item III.2.e).

Endereço: Av. João Naves de Ávila 2121 - Bloco "1A", sala 224 - Campus Sta. Mônica

Bairro: Santa Mônica

CEP: 38.406-144

UF: MG

Município: UBERLÂNDIA

Telefone: (34)3236-4131

Fax: (34)3236-4335

Email: cep@propp.ufu.br



UNIVERSIDADE FEDERAL DE
UBERLÂNDIA/MG



Continuação do Parecer 335.582

UBERLÂNDIA, 21 de Janeiro de 2015

Assinado por:
Sandra Terezinha de Farías Furtado
(Coordenador)

Enderço: Av. João Naves de Ávila 2121 - Bloco "1A", sala 224 - Campus Sta. Mônica
Bairro: Santa Mônica CEP: 38.400-144
UF: MG Município: UBERLÂNDIA
Telefone: (34)3236-4131 Fax: (34)3236-4335 Email: cep@propp.ufu.br