

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
FACULDADE DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

CLÁUDIA LIMA PIMENTEL

**A ELABORAÇÃO DE UM CORPUS ORAL: A ETAPA DE TRANSCRIÇÃO  
DA INTERAÇÃO NA SALA DE AULA DE PORTUGUÊS COMO LÍNGUA  
ADICIONAL**

DISSERTAÇÃO DE MESTRADO

Porto Alegre  
2016

CLÁUDIA LIMA PIMENTEL

**A ELABORAÇÃO DE UM CORPUS ORAL: A ETAPA DE TRANSCRIÇÃO DA  
INTERAÇÃO NA SALA DE AULA DE PORTUGUÊS COMO LÍNGUA  
ADICIONAL**

Dissertação apresentada como requisito final para obtenção do grau de Mestre em Letras pelo Programa de Pós-Graduação da Faculdade de Letras da Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Cristina Becker Lopes-Perna

Porto Alegre  
2016

CLÁUDIA LIMA PIMENTEL

**A ELABORAÇÃO DE UM CORPUS ORAL: A ETAPA DE TRANSCRIÇÃO DA  
INTERAÇÃO NA SALA DE AULA DE PORTUGUÊS COMO LÍNGUA  
ADICIONAL**

Dissertação apresentada como requisito final para obtenção do grau de Mestre em Letras pelo Programa de Pós-Graduação da Faculdade de Letras da Pontifícia Universidade Católica do Rio Grande do Sul.

BANCA EXAMINADORA:

Dr. Cristina Becker Lopes Perna – Presidente (PUCRS)

Dr. Desirée Motta- Motta-Roth (UFSM)

Dr. Ana Cristina Ostermann (UNISINOS)

## AGRADECIMENTOS

Gostaria de agradecer a todos que, de alguma maneira, participaram desta etapa tão importante da minha vida profissional:

À minha orientadora, Prof. Dr. Cristina Becker Lopes Perna, por toda a compreensão, por toda a confiança e parceria durante todas as etapas de desenvolvimento desta pesquisa. Obrigada pela dedicação, pelo carinho, pela amizade e pelas risadas, mesmo nos momentos mais periclitantes.

Ao meu Fã-Clube, pelas risadas, pelas trocas de conhecimento, pelos almoços, congressos, mas, principalmente, pela amizade. Esses dois anos foram muito melhores com a presença de vocês.

À minha família, por toda a compreensão, amor, incentivo e suporte em toda essa caminhada; especialmente aos meus pais, Lorena e Paulo, que sempre me ensinaram a ir em busca dos meus sonhos e apostar nas mais belas formas de se adquirir o conhecimento.

Ao grupo UPLA, por ter aberto as portas para mim, agradeço pelas trocas de conhecimento, pelo engajamento de todos os professores e colegas na busca pelo desenvolvimento de uma nova área de estudo que tem um objetivo muito nobre: estudar para ajudar aprendizes de português no desenvolvimento de uma competência linguística em prol do conhecimento.

A todos os meus amigos que me apoiaram e me incentivaram em todos os momentos dessa caminhada; em especial a minha *guru* do Mestrado, Leticia Presotto, por ter sido a minha guia nesta jornada.

Ao apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (**CAPES**), pela concessão da bolsa que possibilitou a realização deste trabalho.

À Pontifícia Universidade Católica do Rio Grande do Sul, por ter me proporcionado desenvolvimento pessoal e profissional desde a graduação.

*Abra a sua mente, troque o passo, mude de ideia, mas  
jamais perca a sua essência. Uma árvore jamais perde  
suas raízes, não perca as suas.*

(Tati Zanella)

## RESUMO

Os estudos de português como língua adicional (PLA) vêm assumindo uma posição de destaque no cenário brasileiro atual. O UPLA (Uso e Processamento de Língua Adicional) é um projeto da PUCRS, coordenado pela Professora Cristina Becker Lopes Perna, que tem como objetivo desenvolver a competência linguística em PLA de estudantes estrangeiros que estão em mobilidade acadêmica. O projeto conta com trabalhos em línguas como o Mandarim, o Inglês e o Português, com análise em *Metáforas, Hedges e Posicionamento* em âmbito acadêmico - área que estamos descobrindo em função da carência de estudos. Até o momento, os trabalhos têm sido desenvolvidos com base em corpus escrito, pois, no Brasil, o uso de corpora oral é ainda incipiente (PUC-SP, UFMG, PUC-RS, UNISINOS). Inseridos em um contexto específico, que é o da Universidade, a interação dos aprendizes em sala de aula pode fornecer um novo tipo de dado, revelando insights sobre a interação, o uso da linguagem e a aprendizagem. Um corpus de aprendizes de PLA está sendo coletado e elaborado na PUCRS para servir como base para a análise desses dados. Como contribuição para o projeto, este trabalho tem como objetivo apresentar uma proposta de transcrição para representar a interação oral que acontece no ambiente acadêmico geral e mostrar exemplos de análises que poderiam ser feitas a partir dos dados gerados e transcritos da interação em sala de aula de português como língua adicional. A metodologia adotada é a de John DuBois, professor na UC, que criou um sistema de convenções específico para a transcrição da interação. A transcrição da linguagem falada pode fornecer uma ampla gama de informações em muitos aspectos da linguagem, com implicações para a gramática, aspectos semântico-pragmáticos, interação social e da cultura. Esta pesquisa também permitirá o desenvolvimento de materiais para o ensino e o aprendizado do Português Acadêmico, tanto para aprendizes de português como língua adicional como para falantes nativos. Em uma interface com a Pragmática, a Análise da Conversa (AC) e a Linguística de Corpus (LC), a transcrição oral do Português Acadêmico não só irá refletir nossa visão da língua, mas também apresentar padrões de linguagem existentes neste gênero, destacando alguns fenômenos relevantes a serem analisados através do próprio ato de transcrever.

**Palavras-chave:** PLA. Pragmática. Corpus oral. Transcrição. Interação em sala de aula. LC & AC.

## ABSTRACT

Studies on Portuguese as an Additional Language (PAL) have assumed a prominent position in the current Brazilian environment. UPLA (Use and Process of Additional Language) is a project at PUCRS (Pontifical Catholic University of Rio Grande do Sul), located in Brazil, which aims to develop language competence in Academic Portuguese for a growing number of foreign students that are coming to this university to seek higher education. Classroom interaction can provide a new kind of data, revealing insights into the relationship between interaction patterns, language use and learning. Corpora of learners of PAL is being collected at PUCRS and will serve as a basis for the analysis of this data. As a contribution to the project, this paper aims to present a proposal for transcribing the oral interaction in this area and also present interaction analysis models in Portuguese classroom as an additional language. The methodology adopted is that of John DuBois, a professor at the University of California, who created a system of specific symbols for the transcription of the interaction. The transcription of spoken language can provide a wide range of information in many language aspects, with implications for grammar, semantic-pragmatic aspects, social interaction and culture. This survey will also enable the development of materials for teaching and learning Academic Portuguese both for learners of Portuguese as an additional language and for native speakers. In an interface with the Conversation Analysis (CA) and Corpus Linguistics (CL) methodology, the transcription of general academic Portuguese will not only reflect our view of the language, but also present existing language patterns in this genre, highlighting some relevant phenomena to be analyzed through the very act of transcribing.

**Keywords:** PAL. Pragmatics. Oral Corpora. Transcription. Classroom Interaction. CL. CA.

## **LISTA DE SIGLAS**

**AC** – Análise da Conversa

**EAP** – English for Academic Purposes

**IFE**- Inglês para Fins Específicos

**L1**- Língua Nativa

**LA** – Língua Adicional

**LdC**- Linguística de Corpus

**LE** – Língua Estrangeira

**PFA** – Português para Fins Acadêmicos

**PLA** – Português como Língua Adicional

**PUCRS** – Pontifícia Universidade Católica do Rio Grande do Sul

**SL** – Segunda Língua

**UPLA** – Uso e Processamento de Língua Adicional

## LISTA DE ILUSTRAÇÕES

<b>Figura 1 – Lista dos mais de 100 Corpora de Aprendiz do CECL.....</b>	<b>43</b>
<b>Figura 2 – Exemplo de Transcrição.....</b>	<b>56</b>
<b>Figura 3 – Transcrição de um debate em uma aula de PLA.....</b>	<b>64</b>
<b>Figura 4 – Orientação de Trabalho .....</b>	<b>68</b>
<b>Figura 5 – Fotografia .....</b>	<b>69</b>

## SUMÁRIO

<b>INTRODUÇÃO .....</b>	<b>10</b>
<b>1 LÍNGUA ADICIONAL E O PORTUGUÊS PARA FINS ACADÊMICOS .....</b>	<b>14</b>
1.1 O PORTUGUÊS COMO LÍNGUA ADICIONAL.....	14
1.2 O PORTUGUÊS PARA FINS ACADÊMICOS .....	17
<b>2 A PRAGMÁTICA E A INTERAÇÃO .....</b>	<b>25</b>
2.1 A PRAGMÁTICA .....	25
2.2 ANÁLISE DA CONVERSA.....	30
<b>3. LINGUÍSTICA DE CORPUS .....</b>	<b>36</b>
3.1 A LdC.....	36
3.2 CORPUS DE APRENDIZ.....	42
<b>4 A ELABORAÇÃO DO CORPUS E A ETAPA DE TRANSCRIÇÃO .....</b>	<b>48</b>
4.1 CORPUS ORAL DE APRENDIZES DE PLA .....	48
4.2 A TRANSCRIÇÃO .....	50
4.2.1 Definição e características.....	50
4.2.2 O Sistema de Transcrição de Du Bois .....	54
<b>5. TRANSCRIÇÃO E PROPOSTAS DE ANÁLISE DE DADOS.....</b>	<b>58</b>
5.1 A INTERFACE PRAGMÁTICA, AC E LdC. ....	58
5.2 TRANSCRIÇÃO E PROPOSTAS DE ANÁLISES.....	62
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>73</b>
<b>REFERÊNCIAS.....</b>	<b>76</b>
<b>ANEXOS .....</b>	<b>80</b>

## INTRODUÇÃO

Visto que os estudos de Português como Língua Adicional (PLA) vêm assumindo uma posição de destaque no cenário atual, devido à entrada de estrangeiros em diferentes cursos de graduação e pós-graduação nas universidades brasileiras, este trabalho visa contribuir com os estudos do projeto UPLA (Uso e Processamento de Língua Adicional), que tem como objetivo auxiliar o desenvolvimento da competência linguística em PLA de estudantes estrangeiros que estão em mobilidade acadêmica na Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

Trabalhos sobre *Metáforas*, *Hedges* e *Posicionamento* têm sido desenvolvidos em línguas como o Mandarim, o Inglês e o Português em âmbito acadêmico - área que estamos descobrindo em função da carência de estudos. Entretanto, os trabalhos mencionados tomaram como base corpora escrito, pois, no Brasil, é ainda incipiente o uso de corpora oral (UNISINUS, PUC-SP, UFMG, PUC-RS).

Assim como um corpus escrito, um corpus oral exige certos cuidados na sua elaboração, pois deve seguir critérios de representatividade, de etiquetagem e de armazenamento (software adequado para armazenar os dados gerados). Entretanto, na elaboração de um corpus oral, é preciso gravar os dados (áudio ou vídeo), adotar uma transcrição e treinar os transcritores para a realização adequada desse registro, o que demanda tempo.

Pensando em uma iniciativa de criação de acervo e insumo básico para o desenvolvimento de pesquisas na área de português como língua adicional, o grupo caminha na direção do desenvolvimento de pesquisas no âmbito oral, tendo como ponto de partida a construção de um corpus oral de aprendizes de Português como Língua Adicional (PLA) em ambiente acadêmico.

Sendo assim, para dar início a este projeto, esta pesquisa tem como objetivo apresentar uma proposta de transcrição para registrar os dados gerados a partir da interação desses aprendizes na sala de aula de PLA.

Além disso, o trabalho aponta algumas possibilidades, sugere algumas pesquisas que futuramente poderão ser desenvolvidas a partir dos trechos transcritos, tendo como fio condutor a Pragmática, em uma interface com a

Análise da Conversa e a Linguística de Corpus, perspectivas teórico-metodológicas importantes para a investigação de elementos e padrões linguísticos provenientes da fala-em-interação. Nesse sentido, a proposta de análise é ilustrativa, com vistas a caracterizar o potencial de aplicação deste tipo de trabalho.

Como os dados gerados são produções orais resultantes da interação em sala de aula de PLA, a interface entre a Pragmática, a Análise da Conversa (AC), a Linguística de Corpus (LdC) e o gênero acadêmico, denominado de Português para Fins Acadêmicos (PFA) é apresentada como perspectiva teórico-metodológica deste estudo.

A Pragmática (Levinson, 2007) investiga as funções da língua em uso, o que as palavras significam e como se relacionam umas com as outras em determinado contexto. Em outras palavras, ela é responsável pelo estudo dos objetivos da comunicação, fundamental para pesquisas que pretendem explorar o uso da comunicação, os elementos e os padrões linguísticos, por exemplo.

A AC (Kerbrat-Orecchioni, 2006) preocupa-se em analisar detalhadamente como a fala-em-interação é conduzida, em uma atividade por si só e como instrumento para o arranjo da ação social. Essa perspectiva teórico-metodológica visa demonstrar que a conversa não é uma ação caótica e que as pessoas se organizam socialmente através da fala. A combinação da LdC com a AC nos permite analisar itens lexicais em contexto<sup>1</sup> e verificar os padrões da linguagem acadêmica, entretanto, estudos nessa interface só poderão ser desenvolvidos se os dados gerados forem registrados de maneira adequada.

Como proposta para o registro da produção oral dos aprendizes e da elaboração do corpus oral do PFA, apresentamos a transcrição criada por John W. Du Bois<sup>2</sup>. Encontramos na teoria de Du Bois<sup>3</sup> a forma mais coerente de transcrever os dados gerados por nosso grupo (UPLA), pois essa teoria leva em

---

<sup>1</sup> Entende-se por *Contexto* aqui o contexto real, ou seja, a produção do enunciado. “Se o enunciado-objeto está sendo considerado um fragmento retirado de uma conversação, então as propriedades do ato comunicativo devem ser consideradas”. Em COSTA, J. C. **A relevância da pragmática na pragmática da relevância**. Porto Alegre: EDIPUCRS, 2008, p.40.

<sup>2</sup> Professor da Universidade da Califórnia, em Santa Barbara, que possui um projeto chamado *Transcription in Action*, com o objetivo de coletar e disseminar informações sobre a transcrição da fala em interação, incluindo métodos, teorias, ferramentas e pesquisas na área. O site do projeto dá especial atenção à transcrição em linguística, mas também tem como objetivo ser útil para transcritores em outras disciplinas e profissões.

<sup>3</sup> <http://www.linguistics.ucsb.edu/projects/transcription/>

conta a interação, estabelecendo uma interface positiva com as outras perspectivas teórico-metodológicas deste estudo: a AC e LdC.

A necessidade de transcrever a produção oral dos aprendizes de PLA em ambiente acadêmico geral reside no fato de poder, a partir do registro, analisá-la adequadamente. Tornar a fala-em-interação de aprendizes disponível e acessível para estudo possibilitará a pesquisa dos mais variados aspectos linguísticos, dentre eles análises sobre fenômenos como posicionamento, padrões de linguagem, adequação, com vistas à elaboração de materiais para auxiliar os estudantes estrangeiros a desenvolverem sua competência pragmática em PFA. Em outras palavras, o que preocupa o grupo UPLA é o fato dos estudantes estrangeiros aprenderem a estrutura da língua portuguesa, mas não saberem como empregá-la adequadamente, principalmente no ambiente acadêmico, o qual exige desse aluno uma formalidade maior.

Fato é que o aluno aprendiz de português não encontra nos livros didáticos (material ainda inexistente para o ensino de PFA no âmbito da LA) informações sobre a língua em uso. Registrar como os alunos vêm fazendo uso da língua portuguesa no ambiente acadêmico é um pontapé inicial e fundamental para análise de como a interação é estabelecida e de que forma a língua está sendo usada.

Desta forma, este trabalho estrutura-se da seguinte forma: no primeiro capítulo, intitulado *Língua Adicional e o Português para Fins Acadêmicos*, são apresentados e discutidos os conceitos e as definições de Língua Adicional e suas características, bem como a criação de uma área que atenda às necessidades dos alunos em fins específicos, como exige o meio acadêmico, a área de Português para Fins Acadêmicos, que também será relacionada ao estudo.

A *Pragmática e a Interação*, segundo capítulo desta pesquisa, dedica-se à apresentação da Teoria Pragmática, bem como sua importância no desenvolvimento de estudos nesse âmbito. Além disso, traz as definições e conceitos da Análise da Conversa, metodologia que auxilia na análise da interação na sala de aula de PLA. O terceiro capítulo é dedicado à *Linguística de Corpus*, aos tipos existentes, com maior enfoque no tipo de corpora a ser criado pelo UPLA, o corpus de aprendiz.

O quarto capítulo deste trabalho, intitulado *A Elaboração do Corpus Oral e a Etapa de Transcrição*, traz o projeto de elaboração do corpus oral, suas características e etapas em desenvolvimento. Mostra, também, o tipo de transcrição adotada, suas características e a importância do tradutor no processo de registro.

*Transcrição e Propostas de Análise de Dados* compõe o quinto capítulo deste trabalho, apresentando as amostras de transcrição com base no sistema de Du Bois e possibilidades de análise de elementos linguísticos e da língua em uso usando a metodologia da Análise da Conversa em conjunto com a Linguística de Corpus, pois possibilitam um estudo quantitativo e qualitativo ao mesmo tempo. Por fim, são apresentadas as considerações finais do estudo, as possíveis repercussões para a área do PFA e para futuras pesquisas.

## 1 LÍNGUA ADICIONAL E O PORTUGUÊS PARA FINS ACADÊMICOS

Este capítulo subdivide-se em duas seções. Na primeira seção, intitulada *O Português como Língua Adicional*, serão abordados os conceitos de Português como Língua Adicional, contrapondo-o com os termos Língua Estrangeira, Segunda Língua e justificando a escolha do termo em detrimento dos outros. Apresenta-se, também, um breve panorama do PLA nos dias de hoje. Serão discutidas, na segunda seção, *Português para Fins Acadêmicos*, a importância e a necessidade de uma área dedicada ao ensino de português com vistas a aprimorar o ensino e a aprendizagem do estrangeiro que quer aprender essa língua.

### 1.1 O PORTUGUÊS COMO LÍNGUA ADICIONAL

Ao falarmos sobre ensino de línguas, é comum nos confrontarmos com termos como “segunda língua” (SL) e “língua estrangeira” (LE), mas há outro termo que vem sendo bastante utilizado recentemente: o termo Língua Adicional (LA)<sup>4</sup>. Faz-se necessário, entretanto, justificar a escolha de Língua Adicional, e não Língua Estrangeira, adotada neste estudo.

De acordo com Stern (1991) no passado, o termo LE era usado para contrapor o termo “língua nativa” (L1). O termo SL foi incorporado e começou a ser utilizado no ensino de línguas para definir toda e qualquer língua que não fosse classificada como “língua nativa”. O autor explica que, ao falarmos sobre a diferença entre SL e LE, a primeira está relacionada aos falantes que dominam esta língua dentro de fronteiras territoriais onde esta língua é falada como L1 ou “primeira língua”; a segunda, refere-se aos falantes desta língua dentro da comunidade que não apresenta estatuto sociopolítico.

Stern (1991, p.12), apresenta dois significados para o termo SL:

- 1) Refere-se à aprendizagem cronológica da língua, ou seja, qualquer língua que foi adquirida após a “língua nativa”;

---

<sup>4</sup> (Schlater; Garcez, 2009)

- 2) Refere-se ao nível de comando da língua adquirida ao ser comparada à L1. O termo SL está mais relacionado ao nível de proficiência.

Sobre a LE, Stern (1991) diz que ela expressa muito mais a relação que existe entre a língua e a pessoa que a está adquirindo. Em outras palavras, Perna (2016) explica que o aprendiz de uma LE entende que essa língua nunca fará parte do universo linguístico do falante, ao passo que a LA é algo que o falante está acrescentando ao seu universo das línguas nas quais possui competência.<sup>5</sup>

Judd et al. (2001), apresentam a definição de LA no livro intitulado *Teaching Additional Languages*, como qualquer outra língua que esteja sendo adquirida ou aprendida. O uso do termo é mais condizente com aquele estudante que está aprendendo uma língua que não a sua língua nativa, independentemente do lugar onde a aprendizagem esteja acontecendo. Eles argumentam que:

Ensinar uma língua adicional significa ensinar uma segunda, terceira ou outra língua aos alunos que estão dentro dos países de origem ou em países para onde migraram. Justamente porque há tantas línguas no mundo e tantas razões para os estudantes aprendê-las é que o ensino de línguas adicionais torna-se um grande desafio e oportunidade para os educadores (p.03, 2001).<sup>6</sup>

Schlatter e Garcez (2009, p. 37) justificam o uso do termo “LA” reconhecendo que:

Essas Línguas Adicionais são usadas para a comunicação transnacional, isto é, muitas vezes estão a serviço da interlocução entre pessoas de diversas formações socioculturais e nacionalidades, não sendo, portanto, possível distinguir entre nativo e estrangeiro.

---

<sup>5</sup> Segundo comunicação oral da Prof. Dra. Cristina Lopes Perna no 4º Encontro da Rede Sul Letras (12/05/2016,) em Palhoça/SC.

<sup>6</sup> No original: Teaching additional languages means teaching a second, third or further language within students' countries of origin or in countries to which they have migrated. Because there are so many languages in the world and so many reasons why students should learn them, the teaching of additional languages is a great challenge and opportunity for educators. (Tradução nossa).

Com o programa Ciências sem Fronteiras<sup>7</sup>, inaugurado pelo governo com o intuito de fomentar a educação de jovens brasileiros em pesquisas em universidades do exterior, a procura pelo estudo do idioma Inglês teve um aumento significativo. Foi necessário, então, inaugurar simultaneamente o programa Inglês sem Fronteiras. Tendo em vista que a mobilidade acadêmica não apenas acontecia em países de língua inglesa, foi lançado o programa Idiomas sem Fronteiras<sup>8</sup>, oferecendo cursos de francês, espanhol, italiano, japonês, mandarim, alemão e português para estrangeiros.

Entretanto, a mobilidade acadêmica não acontecia apenas em uma direção, considerando o foco global sobre o Brasil em termos socioeconômicos, políticos e culturais. Também houve um aumento significativo na entrada de estrangeiros em instituições brasileiras, principalmente nas de ensino superior, onde as organizações internacionais têm buscado acordos de cooperação com Instituições de Ensino Superior Brasileiras (IES) em todas as áreas acadêmicas.

Como resultado desse intercâmbio, *“o idioma Português está se tornando cada vez mais valorizado como um objeto de atividades educacionais, metodológicas e científicas”* (Santos, 2015, p. 21). Para se ter uma ideia do status do Português hoje, pesquisas indicam ser ela a terceira língua europeia mais falada no mundo, com mais de 250<sup>9</sup> milhões de falantes.

Dessa forma, os estudos de PLA vêm assumindo uma posição de destaque no cenário brasileiro atual, pois há uma procura maior por parte dos estrangeiros que buscam convênio e acordos de cooperação em diferentes áreas acadêmicas nas universidades, necessitando do idioma para sua formação.

Segundo Schlatter e Garcez (2009), as línguas adicionais têm como proposta a promoção do autoconhecimento, da interdisciplinaridade e da participação em contextos e discursos que se organizam a partir da escrita e da fala na língua que está sendo adquirida. Para isso, é necessário que o ensino de PLA seja focado nas reais necessidades desses alunos que adentram as nossas universidades em busca de capacitação.

---

<sup>7</sup>[www.cienciasemfronteiras.gov.br](http://www.cienciasemfronteiras.gov.br)

<sup>8</sup><http://isf.mec.gov.br/>

<sup>9</sup>Fonte: <http://instituto-camoes.pt/lingua-e-cultura/>

Os autores argumentam ainda que só será bem sucedida a educação linguística em Línguas Adicionais se os aprendizes forem capacitados a usar o que aprenderam em sala de aula, no caso deste estudo, de PLA, para poderem participar dos discursos que também se organizam em português, fora do escopo da sala de aula.

Na presente pesquisa, cujo foco é a análise do PLA em ambiente acadêmico para o desenvolvimento da área de PFA (Português para Fins Acadêmicos) para alunos estrangeiros, levantamentos de dados orais dos falantes aprendizes de português em ambiente de sala de aula são extremamente necessários. É a partir desses dados que será possível identificar quais os problemas de produção oral, quais os relacionados à estrutura da interação e os de aspectos sintático-semântico-pragmáticos, que os aprendizes estão enfrentando. Isso nos permitirá, a partir dos dados coletados, elaborar materiais que auxiliem na aprendizagem de PFA, contribuindo para a proficiência do aluno em mobilidade.

É importante desenvolver neste aluno de PLA a proficiência no meio acadêmico, visto que a linguagem utilizada na apresentação de trabalhos e na conversa com o professor e seus pares deve ser adequada ao nível em questão. Uma vez inseridos no ambiente acadêmico, faz-se necessário explorar e definir os conceitos de Linguagem Acadêmica, bem como apresentar a necessidade de se desenvolver uma área de PFA. Esses dois tópicos serão apresentados na próxima seção.

## 1.2 O PORTUGUÊS PARA FINS ACADÊMICOS

No nosso dia a dia nos deparamos com uma variedade enorme de textos e, como falantes nativos, adquirimos essas variedades sem, necessariamente, precisarmos aprendê-las de maneira formal (Biber, 2009). Na escola aprendemos muitas das variedades que não nos são ensinadas em um contexto informal. Aprendemos a ler textos de narrativa, de descrição, e também aprendemos a produzir esses tipos de texto. Conforme o aluno vai progredindo na escola, a demanda vai aumentando e o número de variedades também, até que ele chegue à universidade.

A educação universitária requer a capacidade de ler e compreender a prosa acadêmica, uma variedade que é extremamente diferente da conversa face a face. Além disso, os alunos devem aprender a produzir textos escritos de sub-variedades específicas dentro da prosa acadêmica, incluindo a composição de ensaios persuasivos, relatórios de laboratório dos cursos de ciências, e resumo/síntese em prosa nas respostas dos exames finais. Um dos principais objetivos de uma educação universitária é aprender o registro especializado de uma determinada profissão, seja ela da engenharia elétrica, da química, da sociologia, das finanças, ou da educação Inglesa. O sucesso requer a aprendizagem dos padrões linguísticos particulares que são esperados para situações e finalidades comunicativas específicas (Biber, p. 03, 2009).<sup>10</sup>

Para compreender as palavras de Biber sobre a aprendizagem em um meio específico, faz-se necessário explicar o crescimento e a expansão dos programas voltados à formação desses estudantes.

Em 2014 é lançado o programa Idiomas sem Fronteiras (IsF), destinado à formação e à capacitação de estudantes, professores e corpo técnico-administrativo de instituições de educação superior e de professores de idiomas da rede pública de educação básica<sup>11</sup>. Com o intuito de dar conta desse crescente número de estudantes em mobilidade acadêmica, o programa é voltado aos estudantes e professores que necessitam melhorar a proficiência em línguas como inglês, francês, espanhol, italiano, japonês, mandarim, alemão e português para estrangeiros.

Como estes programas são voltados ao desenvolvimento das línguas em situações do cotidiano, fica a preocupação de não existir um trabalho mais voltado à proficiência em ambiente acadêmico, uma vez que os alunos são requisitados a fazer apresentações de trabalhos e a escrever artigos científicos. É importante

---

<sup>10</sup> No original: A university education requires the ability to read and understand academic prose, a variety that is extremely different from face-to-face conversation. Further, students must learn how to produce written texts from many different specific sub-varieties within academic prose, including persuasive essays in freshman composition, lab reports in science courses, and summary/synthesis prose responses on final exams. One of the main goals of a university education is to learn the specialized register of a particular profession, whether electrical engineering, chemistry, sociology, finance, or English education. Success requires learning the particular language patterns that are expected for particular situations and communicative purposes.

<sup>11</sup> <http://www.brasil.gov.br/educacao/2014/11/novo-programa-oferece-formacao-em-sete-linguas>

que este aluno saiba se expressar, também, de forma culta e adequada ao ambiente acadêmico.

Ao ingressar na universidade, esse aluno estrangeiro, que carrega toda uma bagagem cultural, uma linguagem própria, uma identidade própria, entra em contato com uma nova realidade, uma nova cultura, um meio acadêmico que exige dele uma postura diferente no que diz respeito ao seu posicionamento nos vários tipos de interação.

É preciso, portanto, que o aluno aprenda o “código acadêmico” (Jordan, 1997), que, segundo o autor, envolve adaptação a um novo tipo de sistema, o acadêmico, que possui suas próprias convenções e que pode, também, envolver atitudes e expectativas, algumas das quais expressas pela linguagem.

A linguagem acadêmica, de acordo com Jordan (1997) na definição apresentada em “Inglês para Propósitos Acadêmicos” (EAP, sigla em inglês), pode ser determinada como as habilidades de comunicação requeridas para o estudo em sistemas formais de educação. É o tipo de linguagem utilizada no ambiente acadêmico e para a qual se espera que o aluno tenha competência (Moro, 2014). Esta pode ser tanto escrita (textos utilizados em aula, provas, trabalhos escritos produzidos por alunos, ensaios, teses e artigos) como falada (o discurso do professor, trabalhos orais apresentados por alunos – comunicações individuais, apresentação de pôsteres).

Outrossim, a linguagem acadêmica é aquela produzida no âmbito da universidade, aquela que é usada na apresentação de um trabalho, em uma reunião com um orientador, em um grupo de estudos, na sala de aula para tirar dúvidas ou debater sobre um assunto. Tendo em vista os diferentes âmbitos em que é produzida, essa linguagem pode apresentar diferentes níveis de formalidade, adequados de acordo com o tipo de interação.

No ensino de línguas como o Inglês, existe uma área que atende a essa demanda, a área de Inglês para Fins Específicos (IFE) (Cummins, 2000). É uma área de ensino de línguas que tem como objetivo atender às necessidades do aluno, como, por exemplo, Inglês para Negócios, Inglês Instrumental e Inglês para Fins Acadêmicos, entre outros. A razão para a existência destas especializações é que nem sempre a proficiência em um determinado registro de Inglês irá corresponder a outro tipo de registro nessa mesma língua. Um aluno pode ter um

nível bom de proficiência em conversação, mas não atender aos requisitos da conversação em linguagem padrão, como é o caso da Linguagem Acadêmica, por exemplo.

Soma-se ao fato da necessidade de aprender este novo código linguístico, o fato deste estudante ser um estrangeiro, falante de uma outra língua que não o português, que quer cursar uma faculdade de engenharia de produção (por exemplo) em uma universidade do Brasil. É bem provável que esse aluno tenha estudado e que tenha algumas noções do léxico e da gramática da língua, entretanto, isso não garantirá o seu sucesso no meio acadêmico.

No ensino de PLA não pode ser diferente. É necessário que o processo de ensino e aprendizagem do português seja trabalhado atendendo ao seu fim específico. Os alunos que vêm para a universidade precisam aprender o uso da linguagem em ambiente de ensino superior, ou seja, a Linguagem Acadêmica do Português, ou PFA.

Segundo O'Keefe et al. (2011), estudos em pragmática revelam que a transferência pragmática entre as línguas pode, em certas ocasiões, fazer com que os não-nativos pareçam rudes ou insinceros. Isso mostra o quão importante e necessário se faz o ensino para fins específicos, pois uma vez inseridos em um contexto de PFA, aos aprendizes de PLA deve ser ensinado como fazer uso desse tipo de linguagem em uma forma pragmaticamente adequada.

Segundo Bardovi- Harlig e Mahan- Taylor (2003, p.38):

As conseqüências das diferenças pragmáticas, ao contrário do caso dos erros gramaticais, são muitas vezes interpretadas em nível social ou pessoal e não como resultado do processo de aprendizagem de línguas.<sup>12</sup>

Desse modo, a instrução em pragmática é relevante para prevenir a falha pragmática. Um estudo de Bardovi-Harlig e Dörnyei (1998) examinou alunos e professores de dois países – EUA e Hungria – para testar a adequação de enunciados em diferentes contextos, e mostrou que os aprendizes e professores de inglês como língua estrangeira na Hungria julgaram os erros gramaticais mais

---

<sup>12</sup> No original: The consequences of pragmatic differences, unlike the case of grammatical errors, are often interpreted on a social or personal level rather than a result of the language learning process. (Tradução nossa).

sérios do que os erros pragmáticos, enquanto nos EUA, os aprendizes de inglês como segunda língua e os seus professores julgaram como mais sérios os erros cometidos na esfera pragmática.

Em termos de linguagem de sala de aula (O'Keefe et al., 2011), pode-se dizer que a semântica está relacionada ao ensino do vocabulário, enquanto a sintaxe está relacionada ao ensino de regras gramaticais. Tanto o vocabulário quanto a gramática podem ser ensinados em nível de palavra ou sentença, isolados de qualquer referência ao falante (ou escritor) ou ouvinte (ou leitor). É nessa referência à linguagem dos usuários que reside a separação entre semântica, sintaxe e pragmática.

A competência gramatical e lexical equivale ao conhecimento semântico e sintático do aprendiz na língua alvo. Já a competência pragmática iguala-se ao conjunto de regras internalizadas de como usar a língua adequadamente no âmbito sociocultural, levando em conta os participantes da comunicação e as características do contexto no qual acontece a interação (Celce-Murcia, Olshtain, 2000). O aprendiz de uma língua adicional pode ter o conhecimento gramatical dessa língua adquirida, mas isso não implica que ele saiba como usá-la adequadamente.

Cabe aos professores, portanto, oferecer oportunidades para o desenvolvimento não somente da competência gramatical, mas, também, da competência pragmática, trabalhando a língua em uso em sala de aula. Crandall e Basturkmen (2004) destacam a abordagem inadequada no ensino de Atos de Fala apresentados pela maioria dos livros de ensino de Inglês para Propósitos Acadêmicos. Os autores argumentam que os livros de EAP parecem assumir que os aprendizes já sabem quando é apropriado aplicar os Atos de Fala, ao invés de apresentar quando e em que contexto seus usos podem ser empregados.

Vellenga (2004) aponta que um bom livro didático é aquele que trabalha com atividades que elevam a consciência pragmática, fornecendo aos aprendizes informações contextuais, variedade de formas e informações culturais necessárias para a correta opção pragmlinguística (a gama de formas disponíveis) e a escolha sociopragmática (a forma certa para a situação certa).

Como a área de PFA não está bem estabelecida na literatura, muito do IFE acaba sendo incorporado e adaptado. Entretanto, existem estruturas e aspectos

culturais que cada língua carrega e que não podem ser simplesmente adaptados. Dessa forma, é preciso investigar como os aprendizes usam o português para instruí-los melhor dentro do ambiente acadêmico, tendo em vista que a linguagem que eles utilizam no dia a dia possui um registro mais informal e não atenderia aos padrões exigidos pela academia. Conforme O'Keefe et al.:

Estudos baseados em corpus destacam inconsistências entre o inglês que encontramos nos livros didáticos e o inglês que encontramos em corpus oral ou escrito (2011,p.139) <sup>13</sup>

É nesse sentido que o UPLA visa a contribuir, pois tem por objetivo analisar padrões de linguagem utilizados no PFA via teoria linguística com base em estudos de corpora escrito e oral, tomados a partir de produções de estudantes e professores universitários nativos, bem como a produção de aprendizes, com vistas a oferecer estratégias aproximadas à produção acadêmica da L1.

O estudo baseado em corpora de falantes nativos (Meunier, 2002) é importante porque revela uma descrição detalhada do funcionamento da gramática da L1, características como as frequências das palavras, características sintáticas, combinações e concordâncias da língua fonte são mais precisas quando provenientes de dados empíricos. Assim, é relevante que o grupo trabalhe tanto com corpus nativo quanto com corpus de aprendiz, pois o primeiro traz como base os padrões da língua; o segundo, como esses padrões estão sendo utilizados e de que modo são percebidos pelos aprendizes.

Segundo Meunier (p.123, 2002):

[...] a pesquisa baseada em corpus de aprendiz entra em jogo para ajudar a pesar a importância de cada um dos corpora. A pesquisa baseada em corpus de aprendiz oferece mais refinamento na identificação das formas que são problemáticas para os alunos. Também permite levar em conta a língua materna dos alunos e, conseqüentemente, proporcionar um ensino mais focado e adequado.<sup>14</sup>

---

<sup>13</sup>No original: Corpus studies have highlighted the inconsistencies between the English that is found in textbooks and the English that occurs in spoken and written corpora. (Tradução nossa )

<sup>14</sup> No original: [...] learner corpus research comes into play to help weigh the importance of each of these. Learner corpus research offers further refinement in identifying those forms which are problematic for learners. It also make it possible to take into account the learners' mother tongue and hence provide more focused and appropriate teaching. (Tradução nossa )

Consequentemente, o objetivo é desenvolver práticas educativas nos cursos de PFA com base em pesquisa linguística. Essas práticas vão desde a coleta de corpora oral e escrito tanto de L1 como LA, análise de livros didáticos existentes, desenvolvimento de planos e atividades de aula à elaboração de métodos de avaliação da competência linguística acadêmica de estudantes estrangeiros no Brasil.

Na PUCRS, o grupo UPLA trabalha a partir da necessidade dos alunos em mobilidade acadêmica que chegam na universidade e precisam cursar o português como língua adicional.

A disciplina Português para Estrangeiros (PPE), oferecida pela Faculdade de Letras (FALE) para alunos de graduação e também para alunos de pós-graduação, é um exemplo de políticas educacionais relacionadas ao processo de internacionalização, contribuindo para o fortalecimento da área de Português como Língua Adicional (PLA). A disciplina visa desenvolver as quatro habilidades linguísticas (leitura, escrita, compreensão oral e fala), enfatizando noções fonológicas, gramaticais, discursivas e pragmáticas em língua portuguesa, bem como aspectos culturais, com foco no gênero acadêmico. Propiciar ao aluno condições de se expressar autonomamente, de modo a atingir um nível de competência na compreensão e produção em Língua Portuguesa auxilia o aluno em sua integração acadêmica e profissional, assim como social e cultural.<sup>15</sup>

A criação e o estabelecimento de uma área de pesquisa dedicada ao Português para Fins Acadêmicos (PFA) se justifica na citação acima, pois é imprescindível fornecer a esse aluno condições para que possa integrar-se de maneira acadêmica e social. Para isso, estudos com base em corpora são substanciais para garantir análises profundas de estruturas e padrões da L1 em uma comparação com o que está sendo utilizado como LA, bem como para suplementar a elaboração de materiais didáticos específicos nessa área, que auxiliem na produção acadêmica do estudante de português, que facilitem a sua aprendizagem e contribuam para a formação da sua competência pragmática.

Nesta seção, apresentamos conceitos e definições de PFA no contexto da LA, justificando o uso do termo neste trabalho. Definiu-se, também, o conceito de

---

<sup>15</sup> Ementa do curso de Português como Língua Adicional na PUCRS.

Linguagem Acadêmica e a necessidade de aprofundarmos os estudos para atender aos alunos aprendizes de português que ingressam na universidade e auxiliá-los na condução de seus estudos.

No próximo capítulo, serão apresentadas as perspectivas teórico-metodológicas adotadas para o presente estudo: a Pragmática e a Análise da Conversa.

## 2 A PRAGMÁTICA E A INTERAÇÃO

Neste capítulo serão apresentadas as duas perspectivas teórico-metodológicas para este estudo: a Pragmática e a Análise da Conversa. A primeira seção, intitulada *A Pragmática*, traça um breve histórico da teoria, apresentando suas características e sua importância para os estudos com base nessa perspectiva. A segunda seção, chamada de *A Análise da Conversa*, não só apresenta um breve histórico dessa perspectiva, como mostra sua relevância para os estudos pragmáticos e sua inter-relação com a metodologia de transcrição.

### 2.1 A PRAGMÁTICA

O presente estudo tem a Pragmática como fio condutor. Neste sentido, faz-se necessário, então, definir aspectos importantes dessa abordagem. O uso moderno do termo Pragmática é atribuído ao filósofo pragmático Charles Morris (1938) o qual, com base nas ideias de Peirce, tentou esboçar uma teoria geral dos signos, conhecida como Semiótica (Levinson, 2007). Ele propôs uma tricotomia para a Semiótica, dividindo-a em Sintaxe, Semântica e Pragmática. Nesta divisão, a Sintaxe foi definida como o estudo da relação formal entre os signos, a semântica foi associada às relações dos signos aos objetos e a Pragmática como a relação de signos com os usuários de uma língua. Em outras palavras, de acordo com Levinson (2007, p. 6), a Sintaxe trata do estudo das propriedades combinatórias das palavras e suas partes, a Semântica lida com o significado e a Pragmática estuda o uso linguístico.

Em seu livro intitulado “Pragmática” (2007), Levinson traça um panorama histórico do surgimento da disciplina, apresentando a dificuldade que se tem em definir o termo Pragmática bem como de se delimitar um escopo para essa perspectiva. Segundo o autor, ao tentarmos determinar o estudo da Pragmática, surgem muitas possibilidades de definição, e cada uma delas apresenta deficiências ou dificuldades que acabam não sendo satisfatórias. Levinson dedica muitas páginas de seu livro para mostrar os prós e os contras das inúmeras definições já propostas para a área.

Dentre as definições potenciais mais plausíveis para o termo Pragmática, Levinson (2007, p.25) apresenta essa abordagem como “o estudo das relações entre língua e o contexto que são básicas para a descrição da compreensão da linguagem”, ou seja, entender uma enunciação envolve fazer inferências que irão ligar aquilo que foi dito com o que é suposto ou que fora dito anteriormente. Levinson (2007) menciona também que o estudo da relação entre a língua e o contexto é a base para a descrição da compreensão da linguagem. A compreensão de um enunciado está associada à compreensão das inferências feitas entre o que foi dito com o que foi mutuamente assumido ou com o que foi dito anteriormente. Assim, a Pragmática é responsável pelo estudo das inferências necessárias para a compreensão de um enunciado.

Yule (1996, p4) define a Pragmática como o estudo das relações entre as formas linguísticas e seus usuários. O autor apresenta quatro áreas de estudo da Pragmática, sendo elas definidas como (p.3):

1. O estudo do que o falante quer significar;
2. O estudo do significado contextual;
3. O estudo do quanto pode ser dito além das palavras;
4. O estudo da expressão da distância relativa.

A primeira área está associada ao estudo do significado comunicado pelo falante e interpretado por um ouvinte, ou seja, está mais relacionada àquilo que se quer comunicar do que ao significado literal das palavras ou sentenças proferidas. A segunda está relacionada à importância do contexto nas interpretações das falas dos falantes e o quanto ele pode influenciar o que foi dito.

Na terceira área, o autor menciona que é significativo o estudo do “não dito”, ou seja, aquilo que não foi proferido deve ser reconhecido como parte daquilo que é comunicado. A quarta área de estudo referente à Pragmática aborda o fator que influencia a escolha entre o dito e o não dito. Nesse caso, Yule traz para reflexão o fato de a distância poder determinar o que e o quanto deve ser dito.

A proximidade, seja ela física, social ou conceitual, implica em experiência compartilhada. Os falantes determinam o quanto precisa ser dito partindo do pressuposto da proximidade ou distância do ouvinte (Yule, 1996, p. 3).<sup>16</sup>

Essas áreas de estudo são definidas com o intuito de não restringir a Pragmática apenas ao sentido mais amplo, o da linguagem em uso, mas de tentar apresentar tudo o que ela engloba.

Além de apresentar as áreas de estudo referentes à Pragmática, o autor também fala sobre a importância dos estudos através da abordagem Pragmática. Ele argumenta que a vantagem de se estudar a língua através da pragmática reside no fato de podermos falar sobre a intenção das pessoas ao se comunicarem, o que pretendem, quais os seus objetivos, além de se poder verificar que ações são usadas no momento da fala (Yule, 1996). Esse argumento de Yule é de suma importância para este trabalho, uma vez que temos como um dos objetivos analisar os fenômenos linguísticos resultantes da interação em sala de aula de aprendizes de PLA, justamente para mostrar como a linguagem é utilizada por eles.

Dependendo do contexto em que os termos linguísticos são empregados, eles podem desempenhar diferentes funções. A Teoria dos Atos de Fala contempla essa afirmação, na qual os enunciados podem ser descritos de acordo com as funções que desempenham (O’Keeffe, 2011).

A Teoria dos Atos de Fala surgiu nos anos de 1960 com a publicação de “How to do things with words”, de J. L. Austin. O ponto central desta teoria reside no fato de que o enunciado de uma sentença é parte de uma ação moldada pelas convenções e instituições sociais (Huang, 2007). Nas palavras de Moro (2014, p. 30), “a ideia central da teoria é a de que ao falarmos, já realizamos uma ação (sobre o mundo e sobre o outro)”.

A intenção dos falantes, suas expressões e seus objetivos aparecem de forma muito mais clara em discursos autênticos. É o discurso autêntico, ou seja, uma produção oral e interativa, que possibilita ao pesquisador analisar certas

---

<sup>16</sup> No original: Closeness, whether it is physical, social, or conceptual, implies shared experience. On the assumption of how close or distant the listener is, speakers determine how much needs to be said. (Tradução nossa).

características do discurso falado e até mesmo fenômenos de interesse mais particular (Kasper, Rose, 2002).

Os mesmos autores, ao falarem sobre as abordagens de desenvolvimento em pesquisas pragmáticas mencionam que em pesquisas pragmáticas com o foco em interlínguas, o discurso autêntico pode ser coletado por notas de pesquisa de campo ou gravações de áudio ou vídeo.

Se o objeto de estudo tem foco no contexto discursivo e social, se leva em consideração as tomadas de turno, de perguntas e respostas, ou seja, se leva em conta a interação, então é importante que os dados sejam gravados em vídeo, pois se forem apenas coletados via notas de pesquisa de campo, detalhes importantes da interação podem ser perdidos.

[...] a interação exibe uma prosódia finamente estruturada, detalhes temporais, sequenciais e não-verbais que escapam até mesmo ao observador mais treinado, e que são impossíveis de serem fixados na memória. (Kasper, Rose, 2002, p. 81.)<sup>17</sup>

Para a coleta de produções orais, como é o caso desse estudo, se faz necessário o uso de tecnologias. Os dados aqui gerados são produções orais resultantes da interação dos aprendizes de PLA em sala de aula que foram gravados por uma câmera para que, assim, fosse captada, além do discurso falado, a interação professor- aluno e vice-versa.

Seguindo a linha de estudos através da abordagem Pragmática, encontra-se em Leech (1983) uma definição para essa perspectiva que também vai ao encontro do que se pretende no presente estudo: verificar o tipo de linguagem utilizada por aprendizes de PLA em mobilidade acadêmica. Sendo assim, Leech (1983, p.xi) diz que a Pragmática pode ser definida como o estudo dos significados dos enunciados inseridos em um contexto. Não se pode entender a linguagem sem entender como ela é utilizada na comunicação, e para isso, os estudos de Pragmática são fundamentais.

Ao falar sobre o tipo de linguagem utilizada por esses aprendizes de PLA, torna-se fundamental explicar aqui o termo *Corpus de Aprendiz*, em inglês,

---

<sup>17</sup>No original: [...] interaction exhibits finely structured prosodic, temporal, sequential, and non-verbal details that scape even well-trained observer and are impossible to fix in memory. (Tradução nossa)

*Learner Corpora*, que pode ser definido, de acordo com Granger (2002, p. 7), como “uma coleção eletrônica de dados de alunos aprendizes de segunda língua ou de língua estrangeira”<sup>18</sup>, ou seja, um corpus de aprendiz tem por objetivo descrever a língua dos aprendizes. No entanto, é importante ressaltar que estamos falando de um corpus oral, é a coleta da fala de alunos advindos de outras universidades e que estão aprendendo o PLA. O termo “aprendiz”, nesse caso, está relacionado ao estrangeiro que está aprendendo a língua portuguesa em um ambiente em que ela é a língua nativa.

Torres (2015, p.17) ressalta, em seu trabalho sobre a escrita acadêmica de português por hispano falantes, a importância dos corpora de aprendizes para pesquisas no campo de ensino/aprendizagem:

Esses recursos costumam ser muito úteis em pesquisas dedicadas a conhecer o processo de aprendizagem de uma língua estrangeira, assim como para procurar métodos mais eficientes no ensino/aprendizagem da língua. Seguindo essa linha, os pesquisadores empregam os corpora de aprendizes para estudar questões teóricas ou educacionais, enquanto os educadores se concentram em identificar as necessidades dos aprendizes.

É importante ressaltar a relevância dos estudos sobre desenvolvimento pragmático, pois os estudantes que estão aprendendo o PLA precisam ter consciência sobre alguns aspectos da língua que não são percebidos facilmente.

É justamente através da interação em sala de aula que se pode verificar como a linguagem dos aprendizes de PLA está sendo utilizada, como essa competência está sendo desenvolvida e que tipo de consciência eles têm sobre a língua e, principalmente, sobre o PFA.

Sendo assim, contribuindo para o estudo proposto, a Pragmática se preocupa com os estudos da linguagem considerando o discurso sem isolar a língua da sua produção social. Portanto, os estudos pragmáticos visam definir e analisar o que é a linguagem, trazendo para a sua definição, conceitos de sociedade e de comunicação (Pinto, 2012).

Através dos estudos com base na Pragmática será possível, futuramente, auxiliar no estabelecimento e desenvolvimento de uma área de pesquisa

---

<sup>18</sup> No original: Computer Learner corpora are electronic collections of authentic FL/SL textual data [...]. (Tradução nossa).

dedicada ao PFA, com materiais didáticos específicos nessa área, que auxiliem na produção acadêmica do estudante de português e que facilitem a sua aprendizagem.

Contribuindo para o discernimento dos fenômenos pragmáticos está a Análise da Conversa (AC), que trata da fala-em-interação. Devido a sua relevância para este estudo, visto que os dados gerados neste trabalho são resultados de interação, essa perspectiva teórico-metodológica será abordada na próxima seção.

## 2.2 ANÁLISE DA CONVERSA

A fala-em-interação é central na organização da atividade humana e na vida social. Como a fala é organizada, como é utilizada pelos falantes e qual é o seu papel no processo social são perguntas que podem ser explicadas pela perspectiva teórico-metodológica da Análise da Conversa (AC).

A AC aparece pela primeira vez nos estudos de Harvey Sacks, que nos anos de 1960, na Universidade da Califórnia, pesquisou a organização estrutural da linguagem cotidiana em uso. Influenciado pelos estudos etnometodológicos e pela pesquisa das propriedades estruturais da interação face a face, de Erving Goffman, Sacks inicia um programa de pesquisa destinado a investigar níveis de ordem social que podiam ser revelados através da prática cotidiana da fala (Clift et al., 2009)

A partir daquele momento, Sacks iniciou um estudo sobre a linguagem cotidiana (ordinary talk) com foco nos turnos de fala. Os turnos de fala podem ser definidos, segundo Clift et al. (2009, p.41), como “os métodos pelos quais as pessoas administram a rotina de troca de turnos enquanto minimizam as lacunas e as sobreposições entre eles”.<sup>19</sup>

Sacks, juntamente com Schegloff e Jefferson, foram os primeiros pesquisadores a colocarem os participantes de uma interação no centro da análise (Clift et al., 2009). Isso porque tinham como objetivo descobrir como os

---

<sup>19</sup>No original: [...] the methods by which persons concertedlly manage the routine exchange of turns while minimizing gap and overlap between them. (Tradução Nossa).

participantes compreendem e respondem uns aos outros nos turnos de fala, tendo como foco as sequências de ações que são geradas através dessas produções.

As primeiras observações de Sacks quanto à interação de participantes foram feitas com base nas gravações das ligações das sessões de terapia do *Suicide Preventions Center* (Wooffitt, 2005, p.8). A partir daí, Sacks, Shegloff e Jefferson assumiram que os turnos – enunciados, frases, sentenças e até mesmo palavras isoladas - eram objetos responsáveis por realizar alguma ação ou atividade na interação. Analisar e descrever esses objetos, verificar sua natureza, o que faziam, como ocorriam na interação, como se conectavam ao turno antecedente e quais as suas implicações para os turnos posteriores era o objetivo desses pesquisadores.

A AC pode ser definida, então, como o estudo da conversa, ou seja, o estudo da fala produzida no cotidiano ou o estudo da fala-em-interação. O objetivo dessa perspectiva teórico-metodológica é revelar como o raciocínio é organizado e informado através da fala espontânea e explicar como os enunciados são produzidos e que procedimentos estão envolvidos nessa produção. Hutchby et al. (1998, p.5) definem como objetivo da AC:

[...] descobrir os procedimentos de raciocínio tácito e as competências sociolinguísticas subjacentes à produção e à interpretação da conversa em sequências organizadas de interação.<sup>20</sup>

Ainda sobre o objetivo dessa vertente teórica, Frank e Osterman<sup>21</sup> explicam que:

O objetivo dos estudos em Análise da Conversa é documentar os métodos dos participantes para organizar a vida social. Em última instância, trata-se de observar pessoas agindo em conjunto e produzindo ações sociais para descrever os padrões que elas

---

<sup>20</sup>No original: [...] the objective of the CA is to uncover the tacit reasoning procedures and sociolinguistics competencies underlying the production and interpretation of talk in organized sequences of interaction. (Tradução nossa).

<sup>21</sup> O Capítulo *Pragmática e Análise da Conversa*, de Ingrid Frank (UFRGS) e Ana Cristina Osterman (UNISINOS) IN Perna, Goldnadel e Molsing encontra-se no prelo.

utilizam para construir e manter um entendimento comum acerca do que fazem em conjunto.

Complementando, Kasper e Wagner (2014, p.1) dizem que a AC é mais uma metodologia que auxilia no estudo da interação social e da sua organização sequencial nas mais diversas áreas, como a sociologia, a antropologia, a linguística e a comunicação.

O'Keefe e Walsh (2012) argumentam que essa perspectiva tem como foco as tomadas de turno no processo de interação e procura observar como os participantes estabelecem e mantêm a organização e a sequência da fala para atingir o objetivo da comunicação.

Na Linguística (Hutchby et al., 1998, p. 5), a AC é utilizada por três principais áreas:

- 1) Etnografia da Comunicação –preocupa-se em analisar os padrões da língua em uso e a forma como eles se relacionam com os padrões sociais e culturais;
- 2) Pragmática – procura investigar como o significado é comunicativamente estabelecido;
- 3) Análise do Discurso<sup>22</sup> – ocupa-se com as propriedades estruturais e sequenciais da língua falada.

A AC é baseada na premissa de que o contexto nunca é estático (Walsh, 2013, p. 100). Pelo contrário, ele é formado pelos participantes e pela maneira como eles conduzem a linguagem, como fazem uso dela e, nessa construção, se encaixam as tomadas de turno, as sequências de atos, as aberturas e os fechamentos das conversas, entre outras características. É nesse sentido que a AC contribui para a análise dos fenômenos pragmáticos, pois parte do princípio de que a fala-em-interação acontece em uma situação específica, ou seja, não

---

<sup>22</sup> A Teoria da *Análise do Discurso* surgiu na sociologia do conhecimento científico. Ela tem como objetivo analisar as construções ideológicas presentes em um texto. Esta vertente teórica propõe que a linguagem é usada de forma variável e que os relatos são construídos a partir de possibilidades descritivas que estão ligadas ao contexto em que são produzidas e às funções que realizam (Wooffitt, 2005).

pode ser trabalhada fora de contexto, pois ele é determinante na interação. Entende-se por contexto, aqui, as sequências de ações que acontecem na fala.

Kasper e Wagner (2014, p. 6) apontam alguns procedimentos fundamentais no uso de CA como metodologia de análise:

1) Os dados precisam ser reais, advindos da interação de participantes – os registros gerados precisam ser gravados e devem ser de ocorrência natural, ou seja, sem intervenções por parte de pesquisadores.

2) Os dados precisam ser transcritos e seguidos por anotações – os registros da fala-em-interação precisam ser registrados, ou seja, transcritos de acordo com um sistema que atenda a esses fins.

3) A análise dos dados começa pela sessão de dados, que são discutidos após serem ouvidos e relidos repetidamente – “*As análises se valem das perspectivas e dos entendimentos dos próprios participantes*” (Frank e Osterman)<sup>23</sup>

4) Após a sessão de dados, o pesquisador constrói uma coleção de instâncias comparáveis para, posteriormente, descrever o fenômeno observado – o interesse reside nas ações dos participantes, ou seja, no uso da linguagem.

Frank e Osterman postulam que:

[...] os analistas da conversa interessam-se pelo uso da linguagem (e de outros recursos semióticos que os participantes lançam uso) porque é assim que as pessoas se organizam socialmente para produzir ações no mundo. O interesse, portanto, é no que as pessoas *fazem* umas com as outras e no mundo mediante a troca de turnos de fala. Como vimos anteriormente, a ação é sempre uma troca; ou seja, ela não existe em um turno isolado. Esse, na verdade, é um dos pontos principais que distingue a AC de perspectivas pragmáticas de uso da linguagem. Na perspectiva da AC, não se descrevem promessas, pedidos de ajuda, cumprimentos em turnos isolados. É preciso observar os entendimentos que as pessoas produzem em conjunto sobre as ações que estão desempenhando ao trocarem turnos de fala, pois é assim que elas se organizam socialmente.<sup>24</sup>

<sup>23</sup> O Capítulo *Pragmática e Análise da Conversa*, de Ingrid Frank (UFRGS) e Ana Cristina Osterman (UNISINOS) IN Perna, Goldnadel e Molsing encontra-se no prelo.

<sup>24</sup> No prelo.

Uma maneira de se analisar a língua através da perspectiva da AC é coletar dados provenientes da interação em sala de aula, como é o caso do presente estudo. Segundo O’Keeffe e Walsh (2012), a AC fornece informações sobre detalhes microscópicos da interação, como pausas, tomadas de turno, entonação, sobreposições na fala, entre outros, nos permitindo, assim, analisar a língua em uso através de dados reais, sendo a sala de aula um ambiente ideal para a interação.

Walsh (2013, p.101) aponta como principais características da AC, ao se estudar a interação em sala de aula de uma segunda língua (Língua Adicional no caso deste trabalho) o fato de a estrutura organizacional da interação ser determinada pelos participantes, de forçar o pesquisador a focalizar nos padrões interacionais que emergem dos dados, do contexto ser dinâmico e conduzido pelos participantes, e de enfatizar o contexto e a sequência dos enunciados.

Como o propósito da AC é verificar como os participantes colaboram na construção e na identificação de ações coerentes, devem ser gravadas as interações que ocorrem naturalmente, por isso os dados gerados deste tipo de interação são chamados de dados reais.

As gravações de interações que ocorrem naturalmente são transcritas de modo a captar a produção temporal das declarações em tomadas de turno e, assim, tornar disponível para análise como os participantes a compreendem e respondem uns aos outros (Clift et al., 2009, p.40).<sup>25</sup>

As análises devem ser baseadas em exemplos transcritos das gravações das falas decorrentes de ocorrências naturais da língua (Hutchby et al, 1998). O sistema de transcrição utilizado é seletivo, ou seja, ele não captura todas as possíveis distinções que a gravação pode fornecer. Pelo contrário, as transcrições têm por objetivo fornecer uma interpretação não apenas detalhada, mas acessível, para que os pesquisadores da AC possam analisar os caminhos pelos quais os participantes realizam, de forma ordenada e inteligível, a interação social (Clift et al. ,2009).

---

<sup>25</sup> No original: Recordings of naturally –occurring interaction are transcribed in such a way as to capture the temporal production of utterances in turns-of-talk and thus make available for analysis how participants understand and respond to one another.(Tradução nossa).

É por esse motivo que a coleta de dados e a transcrição são parte fundamental para a AC e para este estudo. É através da transcrição da fala-em-interação que os turnos de fala se tornam passíveis de serem analisados. Isso vai, portanto, ao encontro dos objetivos deste trabalho, que parte do princípio que a transcrição é peça chave na elaboração de um corpus oral, pois torna os dados resultantes da interação da sala de aula de PLA acessíveis para os pesquisadores. Os dados apresentados através dessa transcrição são provenientes da língua em uso, ou seja, são dados reais dos aprendizes de PLA. Além disso, são essas transcrições que irão fomentar análises e conclusões sobre a produção linguística e as dificuldades dos aprendizes com relação à aquisição da língua portuguesa, permitindo, então, uma reflexão sobre o modo como esses fazem uso do português e, ainda, a criação de materiais e estratégias com foco na melhoria desse aprendizado.

Nesta seção, foi apresentada a perspectiva teórico-metodológica da Análise da Conversa, um breve histórico, características e relevância para o estudo desenvolvido. No próximo capítulo, será apresentada a metodologia da LdC, os tipos de corpora existente, o corpus de aprendiz e a relação entre a LdC e a AC, duas vertentes utilizadas em conjunto neste trabalho.

### 3. LINGUÍSTICA DE CORPUS

Apresenta-se, neste capítulo, a metodologia deste estudo. Na primeira seção, com o título de *A LdC*, são apresentados os conceitos dessa metodologia e os tipos de corpora disponíveis para análise de aspectos linguísticos. Na segunda seção, *Corpus de Aprendiziz*, mostra-se os conceitos e as características deste tipo de corpus e a sua relação com o presente trabalho.

#### 3.1 A LdC

Com o surgimento do computador e a posterior popularização e sofisticação de equipamentos, ferramentas e programas, a Linguística de Corpus (LdC) ganhou força, possibilitando o estudo dos mais variados aspectos da linguagem em dados que mostram os usos reais da língua (Pimentel, 2008).

A LdC é uma metodologia em crescente expansão. Ela é definida por Berber Sardinha (2004, p. 03) como uma área que:

[...] ocupa-se da coleta e da exploração de *corpora*, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servir para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração por meio de evidências empíricas, extraídas por computador.

A LdC trabalha com uma abordagem empirista e vê a língua como um sistema probabilístico. Na Linguística, empírico significa “*primazia aos dados provenientes da observação da linguagem, em geral reunidos na forma de um corpus*” (Berber Sardinha, 2004, p.30). Kennedy (1998) ressalta a importância de estudos com base em corpus, pois a LdC não se preocupa apenas com as palavras, estruturas ou usos possíveis da língua, mas também com o que é provável, ou seja, o que é plausível de acontecer na língua em uso. A visão de linguagem como sistema probabilístico significa que “*embora muitos traços sejam possíveis teoricamente, não ocorrem com a mesma frequência*” (Berber Sardinha, 2004, p.31). Para o autor, a frequência é muito importante porque as diferenças entre os corpora não são aleatória, mas motivadas por necessidades

comunicativas nos diferentes gêneros textuais. Essas diferentes necessidades comunicativas, por sua vez, são expressas por diferentes unidades lexicais.

Hunston (2006) define corpus como uma coleção de exemplos naturais da língua, consistindo de pequenas frases até uma quantidade significativa de textos escritos ou fitas gravadas que foram coletados com o intuito de servirem para estudos linguísticos. Um corpus é sempre elaborado para um fim específico. O tamanho e o tipo do corpus dependem do propósito para o qual será usado.

McEnery<sup>26</sup> caracteriza o corpus como uma coleção de palavras armazenadas em computador que têm como intuito servirem a algum propósito. O autor complementa que a LdC não é uma teoria da língua, mas uma metodologia que aborda o estudo da língua.

Biber (p.04, 1996) aponta algumas características das análises baseadas em corpus:

- 1) os dados são empíricos, permitindo a análise de padrões de uso em textos naturais;
- 2) utiliza uma vasta gama de textos naturais para análise;
- 3) faz uso de computador para analisar, utilizando técnicas interativas e automáticas;
- 4) depende de técnicas quantitativas e qualitativas.

O autor ressalta que o importante nas análises baseadas em corpus é ir além das características simples, incluindo análises qualitativas e interpretações funcionais dos padrões quantitativos.

Os registros apresentados em um corpus são exemplos naturais da língua, ou seja, são fragmentos retirados de livros, revistas, jornais, crônicas, incluindo também transcrições da língua falada, que nos permitem analisar o funcionamento da linguagem. Esse armazenamento se dá através de ferramenta computacional e possibilita inúmeros estudos, dentre eles, o de uma Língua Adicional, pois permite que o estudante trabalhe e aprenda através de exemplos e situações reais.

---

<sup>26</sup> Tony McEnery é responsável pelo curso *CORPUS LINGUISTICS: METHOD, ANALYSIS, INTERPRETATION*, da Universidade de LANCASTER. Curso disponível em: <https://www.futurelearn.com/courses/corpus-linguistics/3/steps/36571>

Atualmente, existe um grande número de corpora de diferentes tipos e que atendem a vários propósitos. Os corpora podem ser classificados como: corpus especializado, corpus geral, corpus histórico, corpus monitor, corpus paralelo e corpus de aprendiz (O’Keefe et al., 2011), sendo que o último atende à proposta deste trabalho e será explorado na seção seguinte.

O corpus especializado é aquele em que os dados são característicos de algum tipo de texto ou fala, como é o caso do Michigan Corpus of Academic Spoken English (MICASE) e do British Academic Spoken Corpus (BASE). O primeiro possui uma coleção de transcrições das falas acadêmicas gravadas na Universidade de Michigan<sup>27</sup>; o segundo, transcrições de palestras e seminários gravados em duas universidades do Reino Unido durante o período de 1998 a 2005.<sup>28</sup>

Corpus Geral é aquele que apresenta vários tipos diferentes de textos, como o American National Corpus (ANC)<sup>29</sup>, que contém textos de diferentes gêneros e transcrições da língua falada de 1990 até o presente momento, e o Corpus of Contemporary American English (COCA)<sup>30</sup>, que contém mais de 520 milhões de palavras (20 milhões de palavras por ano 1990-2015) e é dividido igualmente entre linguagem falada, de ficção, de revistas populares, de jornais e textos acadêmicos.

Corpus histórico pode ser definido como aquele que contém textos de diferentes períodos e que nos permite estudar as mudanças que ocorrem na língua. Exemplo disso é o British National Corpus (BNC)<sup>31</sup>, que contém amostras de língua escrita e falada coletadas a partir do final do século XX. Já o corpus monitor tem foco na linguagem atual. O Collings and Birmingham University International Database (COBUILD)<sup>32</sup> pode ser usado como um corpus monitor.

Um corpus paralelo pode ser definido como aquele que possui dois (ou mais) corpora constituídos de pelo menos duas línguas, contendo textos

<sup>27</sup> Michigan Corpus of Academic Spoken English (MICASE) Disponível em: <http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

<sup>28</sup> British Academic Spoken Corpus (BASE). Disponível em: <https://www.sketchengine.co.uk/british-academic-spoken-english-corpus/>

<sup>29</sup> American National Corpus (ANC). Disponível em: <http://www.anc.org/>

<sup>30</sup> Davies, Mark. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Disponível em: <http://corpus.byu.edu/coca/>.

<sup>31</sup> Davies, Mark. (2004-) *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Disponível em: <http://corpus.byu.edu/bnc/>.

<sup>32</sup> <http://www.collins.co.uk/category/English+Language+Teaching/COBUILD+Reference/>

traduzidos de uma língua para a outra (uma obra traduzida do inglês para o português, por exemplo) ou textos produzidos simultaneamente em duas ou mais línguas. Um exemplo de corpus paralelo é o COMPARA<sup>33</sup>, um corpus bidirecional de português e inglês, ou seja, uma base de dados com textos originais nestas duas línguas e as suas respectivas traduções, ligadas frase a frase.

O corpus de aprendiz é aquele que possui dados de linguagem usada por pessoas que estão aprendendo uma língua adicional<sup>34</sup>. Com o corpus de aprendiz uma análise das interferências da língua nativa no processo de aquisição de uma língua adicional pode ser realizada, permitindo um estudo mais aprofundado sobre o processo (Torres et al., 2014). Um exemplo de corpus de aprendiz é o ALEC (The Advanced Learner English Corpus), da Universidade de Uppsala. É um corpus escrito formado por composições de estudantes universitários da linguística e da literatura inglesa, que possui 1,3 milhões de palavras e conta com mais de 146 textos.<sup>35</sup>

É notório o desenvolvimento desta área em trabalhos que envolvem a língua inglesa e são inúmeros os exemplos de corpora, dos mais variados tipos existentes disponíveis para estudo dessa língua. Sardinha (2004) diz que no Brasil, as pesquisas baseadas em corpora são mais desenvolvidas por grupos que trabalham com o processamento da linguagem natural, através da Linguística Computacional e da Lexicografia. Além disso, a maior parte dos corpora elaborados até agora são escritos. Abaixo, encontram-se alguns exemplos de corpora na língua portuguesa:

- *Corpus do Português*<sup>36</sup> – constituído de mais de 45 milhões de palavras de quase 57.000 textos em português do século XIV ao XX. O corpus está dividido de três maneiras diferentes: Registro (comparação entre o oral, a ficção, o jornalístico e o acadêmico),

<sup>33</sup><http://www.linguateca.pt/COMPARA/>

<sup>34</sup>Explicação dada por Tony McEnery no curso *CORPUS LINGUISTICS: METHOD, ANALYSIS, INTERPRETATION*, da Universidade de LANCASTER. Curso disponível em: <https://www.futurelearn.com/courses/corpus-linguistics/3/steps/36571> . Acesso em julho de 2016.

<sup>35</sup><http://www.helsinki.fi/varieng/CoRD/corpora/ALEC/>

<sup>36</sup>Davies, Mark and Michael Ferreira.(2006-) *Corpus do Português: 45 million words, 1300s-1900s*. Disponível em: <http://www.corpusdoportugues.org>.

Dialeto (português brasileiro x europeu no século XX) e Período Histórico (comparação dos séculos diferentes do XIV ao XX).

- *C-ORAL - BRASIL*<sup>37</sup> – Corpus de Referência do Português Brasileiro – é um corpus de referência do português falado informal, organizado por Tommaso Raso e Heliana Mello (ambos da Universidade Federal de Minas Gerais, UFMG). O objetivo desse corpus é compor com o C-ORAL-ROM, o corpus europeu das 4 principais línguas românicas europeias.
- O *FUTLEX* (Perna, Delgado, Todt, et al, 2013) é um corpus multilíngue (inglês – espanhol – português) que apresenta as terminologias específicas da Federação Internacional de Futebol (FIFA). O seu objetivo é estabelecer uma maior padronização e adequação da terminologia utilizada no âmbito do Futebol, entre os meios informativos (rádio e televisão) e entre profissionais que socializam este tipo de conhecimento (tradutores, intérpretes, revisores), visando facilitar a comunicação entre estes profissionais e o entendimento do público leigo nestes três idiomas. Além disso, apresenta como proposta a construção de uma metodologia específica para elaboração deste produto terminográfico, adaptada à necessidade profissional determinada pelo mercado de trabalho e à realidade cultural do país.
- *COMPARA*<sup>38</sup>, um corpus bidirecional de português e inglês, é uma ferramenta que permite estudar a tradução humana e contrastar o português e o inglês através de pesquisas automáticas. Por exemplo, se inserirmos uma palavra em português, podemos ver como essa palavra foi traduzida para inglês em diferentes contextos.

---

<sup>37</sup> Projeto C- ORAL-BRASIL I, de Tommaso Raso e Heliana Mello (UFMG). Disponível em: <http://www.c-oral-brasil.org/>

<sup>38</sup><http://www.linguateca.pt/COMPARA/>

Os corpora têm uma larga escala de usos, podendo ser usados para o ensino de línguas, visto que fornecem informações sobre o funcionamento da língua, sobre padrões discursivos, sobre posicionamentos e hedges; para observarmos diferenças entre uma língua e outra; auxiliar tradutores, pois permitem que verifiquem como as palavras são traduzidas de uma língua para outra e servem, ainda, para estabelecer normas de frequência no uso da língua, e para a investigação de atitudes culturais expressas através da linguagem (Pimentel, 2008).

Kennedy (1998) argumenta que as análises feitas a partir de corpora não só podem contribuir para a construção de dicionários, listas de palavras, gramáticas descritivas, como também para estudos estilísticos, pedagógicos, entre outras implicações.

No mesmo sentido, Pacheco (2010, p.81) complementa:

A LdC torna possível um tipo de descrição linguística que pode ser muito proveitosa para os profissionais do ensino de língua estrangeira, já que esses terão acesso a informações sobre os estágios de aquisição de determinadas estruturas como língua estrangeira de forma sistematizada, ou seja, organizadas de maneira a poderem ser investigadas eletronicamente.

Apesar dos avanços da LdC no Brasil, estudos voltados para a aquisição de PLA são ainda incipientes e são discretas as ações para a criação de um corpus de aprendizes nessa área. Entretanto, precisamos dessas fontes (corpus de aprendiz) para desenvolver mais pesquisas na área, para estudar como os falantes estão fazendo uso das línguas e de que forma podemos criar estratégias para desenvolver a competência deles na língua portuguesa.

Na próxima seção serão apresentadas as definições de corpus de aprendiz e suas características, bem como os corpora existentes nessa área e a justificativa da escolha deste tipo de corpus para a presente pesquisa.

### 3.2 CORPUS DE APRENDIZ

Como já foi mencionado, um dos objetivos do grupo UPLA é, após definida a metodologia de transcrição, elaborar um corpus oral de aprendizes do PFA. Para tanto, faz-se necessário definir melhor o que é e qual a finalidade desse tipo de corpus, uma vez que as interações gravadas pelo grupo são de estudantes advindos de outras universidades no exterior e que estão aprendendo o português como LA.

Um corpus de aprendiz pode ser definido , segundo O’Keefe et al. (2011,p.7), como:

[...] coleções de textos produzidos por aprendizes de uma língua. Eles permitem ao pesquisador identificar padrões em uma variedade particular da linguagem do aprendiz e compará-la a de outros usuários de uma língua.<sup>39</sup>

Neste estudo, porém, as coleções armazenadas não serão escritas, mas orais. Sendo assim, existem algumas características mais específicas para a definição de corpus oral de aprendiz. Granger (2002), por exemplo, menciona que para ser classificado como um corpus de aprendiz, as amostras de discurso apresentadas precisam ser longas, não podem ser apenas palavras ou sentenças. A autora também fala sobre a questão da autenticidade dos textos e falas produzidos por estudantes de língua adicional, pois o contexto do ensino de línguas sempre envolve certo grau de “artificialidade”, e que os dados gerados das produções desses alunos jamais serão completamente naturais.

Entretanto, Granger (2002, p. 5) argumenta a importância dos estudos com base em corpus de aprendiz, dizendo que:

O Corpus de aprendiz fornece um novo tipo de dado que pode auxiliar tanto na pesquisa em ASL (Aquisição de Segunda Língua), que tenta compreender os mecanismos de aquisição de uma segunda língua ou língua estrangeira, quanto na pesquisa em ESL (Ensino de Segunda Língua), que tem como objetivo

---

<sup>39</sup> No original: Learner corpora contain collection of texts produced by the learners of a language. They allow the researcher to identify patterns in a particular variety of learner language and to compare the language of the learner to that of others users of a language. (Tradução nossa).

aprimorar o ensino e a aprendizagem da segunda língua ou da língua estrangeira.<sup>40</sup>

No mesmo sentido, O'Keefe et al. (2011) também ressalta a importância desse tipo de corpus, uma vez que ele permite ao pesquisador não apenas identificar certos aspectos na aprendizagem de língua, como também destacar quais são as dificuldades do aprendiz. Além disso, o corpus de aprendiz nos permite descrever diferentes variedades que emergem da comunicação entre os falantes nos quais a primeira língua não é a que eles usam para se comunicar.

O CECL (Center for English Corpus Linguistics)<sup>41</sup> é um projeto coordenado pela profa. Dra. Sylviane Granger especializado no uso e na coleta de corpora para fins linguísticos e pedagógicos. Tem como principais áreas de estudo os corpora multilíngue e os de aprendiz.

**Figura 1 – Lista dos mais de 100 Corpora de Aprendiz do CECL**

The screenshot shows the website for the Centre for English Corpus Linguistics (CECL) at the University of Louvain. The page is titled "Learner corpora around the world" and contains a table listing various learner corpora. The table has columns for Corpus, Target language, First language, Medium, Text type/task type, Proficiency level, Size in words, Project director, and Availability.

Corpus	Target language	First language	Medium	Text type/task type	Proficiency level	Size in words	Project director	Availability
The Arabic Learner Corpus (ALC)	Arabic	66 languages	written and spoken	Narrative and discussion	Intermediate and advanced	written: c. 283,000 audio: c. 3h30	Abdullah Alfaihi & Eric Atwell	Available
The Pilot Arabic Learner Corpus	Arabic	English	written	Narrative	Intermediate and advanced	c. 9,000	Ghazi Abuhsakma Reem Faraj Anna Feldman Eileen Fitzpatrick Montclair State University, USA	
The Jinan Chinese Learner Corpus (JCLC)	Chinese	50 languages	written	Exams and assignments	Beginners, intermediate and advanced	c. 6 m. Chinese characters c. 9,000 texts	Maolin Wang Shervin Malmasi Mingxuan Huang	
The AKCES/CZESL corpus (Acquisition corpora of)	Czech	Various	written and spoken	Student essays and interviews	Various	2 m.	Karel Sebesta Charles University in Prague Technical University in Liberec, Czech Republic	Available

Fonte: <http://www.uclouvain.be/en-cecl-lcworld.html>

<sup>40</sup> No original: Learner corpora provide new type of data which can inform thinking both in SLA (Second Language Acquisition) research, which tries to understand the mechanisms of foreign/second language acquisition, and in FLT (Foreign Language Teaching) research, the aim of which is to improve the learning and teaching of foreign/second languages. (Tradução nossa)

<sup>41</sup> <http://www.uclouvain.be/en-cecl.html>

Dentre os corpora de aprendizes existentes atualmente, conforme lista criada por Granger, destaca-se o CLC (Cambridge Learner Corpus of Written English) que conta com 50 milhões de palavras coletadas de provas escritas feitas por estudantes do mundo todo que realizaram os exames da Cambridge<sup>42</sup>.

Cabe ressaltar aqui que, mesmo sendo corpora acadêmicos, a maioria dos corpora apresentados na lista é escrito e não oral:

- O BAWE (British Academic Written English) – corpus escrito que contém dados produzidos por falantes de L2, da Universidade de Birmingham, UK. Acesso gratuito.
- O CALE (Corpus of Academic Learner English) – Corpus escrito, com textos acadêmicos de diferentes tipos. Universidade de Bremen, Alemanha.
- O ACAW (Aachen Corpus of Academic Writing) – Corpus de estudantes alemães com textos escritos em inglês. Escrita de pesquisas acadêmicas.
- O CAWE (The Chinese Academic Written English Corpus) – dissertações escritas por chineses graduando em inglês, linguística e linguística aplicada.
- O LANCAWE (Lancaster Corpus of Academic Written English) – Testes de escrita do IELTS (tarefas descritivas e argumentativas); Dados longitudinais.

No Brasil, destaque para o BR-ICLE, um corpus que integra um projeto mais amplo, o The ICLE, sediado na Bélgica e dirigido por Granger, especialista em Corpora de Inglês de Aprendizes ("Learner Corpora"). No país, o grupo é coordenado pelo prof. Dr. Tony Berber Sardinha da PUC-SP, e tem como objetivo coletar, organizar e informatizar o sub-corpus referente à produção em língua inglesa de alunos brasileiros (falantes, portanto, de Português), ou seja os "Brazilian Portuguese Learners".<sup>43</sup>

---

<sup>42</sup> Cambridge Learner Corpus. Disponível em:

<http://www.cambridge.org/us/cambridgeenglish/about-cambridge-english/cambridge-english-corpus>

<sup>43</sup><http://www2.lael.pucsp.br/corpora/bricle/projeto.htm>

Na Universidade de São Paulo foi criado o COMAprend<sup>44</sup>, um corpus de aprendizes formado por redações dos aprendizes de línguas estrangeiras de cinco áreas do Departamento de Letras Modernas das Faculdades de Filosofia, Letras e Ciências Humanas - alemão, espanhol, francês, inglês e italiano. Inclui material produzido nos cursos de graduação e nos seus cursos extracurriculares. É considerado, também, um corpus multilíngue, pois apresenta a nacionalidade brasileira como comum aos aprendizes, nas cinco línguas-alvo.

Ainda na Universidade de São Paulo, o COMET <sup>45</sup>(Corpus Multilíngue para Ensino e Tradução), é um projeto coordenado por Stella Tagnin em parceria com o BrICLE e tem como objetivo servir de suporte para pesquisas linguísticas, principalmente as relacionadas à tradução, terminologia e ensino de línguas.

Torres (2015) e Rojas (2006) são exemplos de pesquisadores brasileiros que criaram corpora de aprendizes para avaliar o desempenho de hispano falantes durante a aquisição do PLA. Rojas (2006), em sua pesquisa, tinha como objetivo apresentar a natureza dos erros produzidos por hispano falantes e, para isso, ele criou um corpus oral com trechos de entrevistas de seis hispano falantes. O pesquisador fez um levantamento das características dos erros produzidos durante diferentes estágios de aquisição. Torres (2015) propôs a criação de recursos e ferramentas de suporte para a escrita no nível lexical como forma de aprimorar a qualidade linguística de textos produzidos por hispano falantes na língua portuguesa. A autora compilou léxicos bilíngues a partir de um Corpus Paralelo de português – espanhol disponibilizado por Azis e Specia (2011) que contém textos provenientes de uma edição on-line da Revista Pesquisa FAPESP.

Nesta perspectiva, Pacheco (2010), em seu estudo sobre a aquisição de morfemas em inglês como L2 por aprendizes falantes de português brasileiro como L1, desenvolveu um corpus de aprendiz para análise desses aspectos linguísticos. O BELC – *Brazilian English Learner Corpus* – possui textos produzidos por alunos (graduados ou graduandos de diversas áreas) do curso de inglês geral oferecido pela PUCRS que adotaram as aulas como disciplina eletiva. O BELC conta com textos produzidos por 424 sujeitos, totalizando aproximadamente 103 mil palavras.

---

<sup>44</sup><http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/11493/6770>

<sup>45</sup> <http://comet.fflch.usp.br/projeto>

Entretanto, não se encontra nenhum estudo baseado em produção oral de aprendizes de PLA, nem mesmo voltados ao registro da produção oral de aprendizes do português em meio acadêmico. Diferentemente do inglês, em que o número de recursos e materiais para o ensino da língua é abundante, sem falar do número crescente de pesquisas que servem para auxiliar o aprendiz de LA, as pesquisas baseadas em corpus com o intuito de gerarem contribuições para a aquisição de PLA andam muito atrás do esperado.

A escolha da elaboração de um Corpus de Aprendiz em detrimento de outros se deu pelo fato da possibilidade de tornar disponível um material rico para análise de padrões linguísticos, de língua em uso, de adequação e competência pragmática, de competência interacional e de pesquisas que possam contribuir com a criação de materiais para aprimorar e enriquecer a aprendizagem do aluno de português como LA.

O estudo baseado em corpora de falantes nativos fornece uma descrição mais precisa das estruturas gramaticais e sintáticas da língua. Entretanto, o foco no estudo de frequências das palavras ou de combinações de nada adianta se não for implementado em atividades que façam parte do currículo (Meunier, 2002). Dentro de um programa de ensino de segunda língua ou de língua adicional, como é o caso deste estudo, é importante que exista um equilíbrio entre a frequência, a dificuldade e a relevância dos dados. Por isso os corpora de aprendiz são tão importantes, porque eles oferecem maior refinamento no estudo da L1. Como argumenta Granger (p.123, 2002), “*quando há um choque entre os insights derivados de corpora de nativos e de aprendizes, é sempre o ângulo do aprendiz que deve ser dada prioridade*”<sup>46</sup>. Em outras palavras:

Estudos baseados em corpus de nativos e de aprendizes estão fornecendo melhores descrições dessas gramáticas. Embora as descrições da língua nativa possam ser tipicamente usadas para definir a agenda de ensino, as descrições dos aprendizes podem ser usadas para modificar esta agenda, levando em conta as necessidades específicas dessas populações (Meunier, p.125, 2002).<sup>47</sup>

---

<sup>46</sup> No original: When there is a clash between the insights derived from native and learner corpora, it is always the learner angle which should be given priority. (Tradução nossa)

<sup>47</sup> No original: Corpus studies based on native and learner corpora are providing better descriptions of native and learner grammars. Whilst native language descriptions could typically be used to define the teaching agenda, the learner descriptions can be used to modify this agenda to take into account the needs of specific learner populations. (Tradução nossa).

Os benefícios da adoção do trabalho baseado em corpora não se limitam apenas à gramática, mas estendem-se para a elaboração de dicionários modernos e livros didáticos. Um dos exemplos é o Collins Cobuild, que tem suas gramáticas baseadas em análises do Bank of English (Meunier, 2002). A gramática do Inglês falado e escrito (Grammar of Spoken and Written English) é a primeira a adotar uma abordagem baseada em corpus para descrever características gramaticais presentes nas diversas variedades do inglês.

No Brasil, a equipe de Tony Berber Sardinha prepara materiais didáticos para o aprendizado de inglês baseados em tarefas que utilizam o Word Smith Tools (Scott, 1997), um programa de análise de corpora.

No próximo capítulo será apresentada a combinação da metodologia adotada neste estudo (a LdC) com a base teórica proposta pela AC e suas contribuições para estudos baseados em dados reais obtidos através da fala – em – interação.

## 4 A ELABORAÇÃO DO CORPUS E A ETAPA DE TRANSCRIÇÃO

### 4.1 CORPUS ORAL DE APRENDIZES DE PLA

Embora o projeto de elaboração do corpus oral de aprendizes de PLA seja bastante incipiente, o corpus a ser criado é de aprendiz, é oral e é acadêmico. O objetivo do grupo UPLA, que trabalha com o uso e processamento de língua adicional, é construir um corpus oral com base nas gravações das interações dos aprendizes em sala de aula de português.

Diferentemente do C-Oral (UFMG), que apresenta um corpus de referência da fala informal do português, nosso objetivo, enquanto grupo, é transcrever o modo como os estrangeiros utilizam a língua, ou seja, nossa preocupação não concerne a descrição fonética da língua, mas está voltada à questão pragmática, onde o uso dos léxicos e das expressões é mais importante.

Sabe-se que um corpus é essencial na pesquisa sobre linguagem em uso, pois trabalha com exemplos naturais da língua, entretanto, ao contrário da elaboração de um corpus escrito, um corpus oral leva mais tempo para ser construído, uma vez que a fala precisa ser transcrita e provavelmente codificada devido a fatores não verbais (O'Keefe et al., 2007).

Um corpus não é apenas uma coleção de dados; ele deve representar uma linguagem ou parte dela (Biber, 1996). A questão da representatividade é muito importante na elaboração de um corpus, o seu design depende muito daquilo que se quer representar. A representatividade de um corpus determina que tipo de pesquisas podem ser feitas e que tipo de perguntas podem ser respondidas através dele.

Todo o processo de elaboração de um corpus, seja qual for o seu propósito, precisa atender a certos quesitos, como qual é o tipo de corpus, para que fim será usado, como será transcrito, etc. Como já mencionado anteriormente, o corpus a ser elaborado é oral, ou seja, os dados gerados são provenientes da fala e não da escrita.

Outra característica do corpus a ser construído é que, dentre os mais variados tipos, ele será um *corpus de aprendiz*, em que as informações coletadas

são representativas da fala de alunos estrangeiros que estão aprendendo o PLA em âmbito acadêmico.

Quando os alunos entram na PUCRS, independentemente do tipo de área e de curso em que estejam inseridos, todos eles precisam fazer aulas de português. A disciplina Português para Estrangeiros (PPE), oferecida pela Faculdade de Letras (FALE) visa desenvolver as quatro habilidades linguísticas (leitura, escrita, compreensão oral e fala), enfatizando noções fonológicas, gramaticais, discursivas e pragmáticas em língua portuguesa, bem como aspectos culturais, com foco no gênero acadêmico.

Propiciar ao aluno condições de se expressar autonomamente, de modo a atingir um nível de competência na compreensão e produção em Língua Portuguesa auxilia o aluno em sua integração acadêmica e profissional, assim como social e cultural.

Para obter os dados da fala-em-interação de PLA, os aprendizes foram filmados em diferentes momentos de aula: em debates, em apresentações de trabalho e em orientações entre aluno e professor orientador acerca de trabalhos. A filmagem dessas interações foi realizada por uma das integrantes do grupo UPLA sem que houvesse qualquer tipo de interferência. A câmera era ligada na sala de aula, em um ponto estratégico, ou seja, em um lugar da sala de aula que pudesse captar o som e as imagem com qualidade, mas que, ao mesmo tempo, não intimidasse os participantes durante a interação, e só era desligada ao final da aula.

Deppermann e Schütte (2008, p. 179) apontam algumas vantagens decisivas da filmagem:

- Não são seletivas, pois mostram todos os detalhes do processo comunicativo;
- Não são afetadas pelos pesquisadores;
- Preservam a qualidade transitória, temporal da interação interpessoal e documentam-na de forma natural;
- Não envolvem uma interpretação completamente ligada aos dados, mas sim fornecem um registro físico e passivo de eventos acústicos e visuais.

Consequentemente, o processo de análise desse tipo de dado pode ser feito em etapas separadas.

Por isso, os dados gerados são de fala espontânea e natural. Sendo assim, este corpus apresentará o registro da interação dos aprendizes no âmbito acadêmico. Para realizar este registro, foi adotada a transcrição de Du Bois, que propõem um sistema de convenções para o registro da fala em interação. Este sistema será melhor abordado na próxima seção desta pesquisa.

Como foi dito anteriormente, os dados serão registrados de acordo com o sistema de convenções de Du Bois, e armazenados, posteriormente, em um software adequado para análise.<sup>48</sup>

Na próxima seção, será apresentada a transcrição adotada para a elaboração do corpus pretendido, bem como suas características e fundamentos.

## 4.2 A TRANSCRIÇÃO

Na seção anterior foi apresentado o projeto de elaboração do corpus oral de aprendizes de português como LA. Sabe-se, entretanto, que essa elaboração exige uma demanda maior. O objetivo da pesquisa é apresentar uma proposta de sistema de transcrição para registrar os dados já gerados para torná-los, assim, possíveis de análise.

Nas próximas seções, serão apresentadas a definição de transcrição, o sistema e as convenções de Du Bois.

### 4.2.1 Definição e características

A transcrição do discurso pode ser definida (Du Bois et al., 2014 ) como um processo em que uma representação gráfica é criada para registrar um evento de fala, tornando-o, assim, possível para análise. O autor complementa:

---

<sup>48</sup> Como a elaboração do Corpus Oral de Aprendizes de PLA está ainda na fase de discussão e desenvolvimento, o software adequado para o armazenamento dos dados gerados pelos alunos de PLA ainda não foi escolhido, mas alguns já estão sendo testados, dentre eles o WordSmith Tools.

O processo de transcrição do discurso nunca é mecânico, mas crucialmente dependente de interpretação dentro de um quadro teórico de referência para se chegar a categorias funcionalmente significantes ao invés de dados acústicos aleatórios.<sup>49</sup>

Para Deppermann e Shütte (1998, p.179), “a transcrição é uma representação de uma gravação de áudio ou vídeo que pode ser editada usando um editor específico para isso”. Os autores complementam que o editor é configurado de acordo com convenções de transcrição especializadas que definem como os fenômenos acústicos ou visuais devem ser processados para o modo de escrita. Além disso, que o editor deve capturar todas as atividades vocais dos participantes da conversa, de acordo com a gravação de áudio ou vídeo.

A transcrição está se tornando uma ferramenta muito importante para muitos pesquisadores, principalmente para os linguistas. Du Bois et al. (2014) apresentam uma vasta gama de áreas que se beneficiam dessa metodologia:

- 1) Gramáticos e linguistas em geral usam as transcrições como fontes de dados linguísticos e para acompanhar a ação em teorias fundamentadas no discurso;
- 2) Linguistas computacionais fazem uso da metodologia para testar protocolos de reconhecimento de voz contra o uso de linguagem real;
- 3) Professores de línguas usam a transcrição para ilustrar os usos reais da língua falada;
- 4) Cientistas sociais usam para compreender a natureza da interação social;
- 5) Alunos de qualquer área podem usar transcrições para saber mais sobre sua área de estudo.

Uma transcrição é realizada para que os pesquisadores possam lê-la. Como os pesquisadores fazem a leitura dessas transcrições, é importante que elas apresentem não apenas o que é necessário, mas que sejam, também, de

---

<sup>49</sup> No original: The process of discourse transcription is never mechanical, but crucially relies on interpretation within a theoretical frame of reference to arrive at functionally significant categories, rather than raw acoustic facts. (Tradução nossa)

fácil compreensão. Através das transcrições, pode-se pesquisar um fenômeno linguístico, verificar um padrão linguístico ou até mesmo formular uma hipótese. Um bom sistema de transcrição deve ser capaz de acomodar, com alguma adaptação, se for necessário, uma gama toda de eventos susceptíveis de serem de interesse do pesquisador. Além disso, o sistema de transcrição adotado deve ser flexível e acomodar necessidades de todos os tipos (Du Bois et al., 2014).

Os registros da fala-em-interação são realizados com o intuito de permitirem ao pesquisador a análise de fenômenos ou aspectos de seu interesse. Para tanto, alguns critérios devem ser obedecidos para uma boa transcrição, como por exemplo, apresentar detalhes da conversa que podem ser relevantes para mostrar como os participantes realizam as suas atividades.

Deppermann e Shutte (1998) apresentam os fenômenos que podem ser representados em uma transcrição:

- 1) Enunciados que contêm detalhes morfo-fonológicos e morfossintáticos, como contrações e variações dialetais de palavras;
- 2) Enunciados vocais, como risadas, assovios, interjeições, hesitações, etc.;
- 3) Turno de fala dos falantes;
- 4) Prosódia;
- 5) Gestos, expressões faciais, postura corporal, movimentos dos falantes, entre outros;
- 6) Tradução (code-switching) – transcrições dos segmentos de fala produzidos em língua estrangeira.
- 7) Enunciados que expressam características individuais da fala, como ironia, surpresa, susto, etc.

A transcrição (Du Bois et al., 2014) descreve e documenta o uso da linguagem. Apesar de o discurso escrito ser mais fácil de ser coletado e analisado e não necessitar de transcrição, ele se difere da fala em estrutura. O registro da fala, por exigir grande esforço (precisa ser transcrito) não vem sendo documentado.

Esse registro da fala-em-interação é de suma importância, pois através da transcrição podemos analisar aspectos relacionados à gramática, ao significado e ao léxico. É na fala que o processo de produção de linguagem se torna mais evidente para o pesquisador. As hesitações, as pausas, os turnos de fala, entre tantos outros fenômenos são observados na produção oral.

São esses fenômenos que oferecem pistas sobre os participantes, o modo como eles organizam o discurso, como produzem seus enunciados, como participam e constroem a interação. Além disso, a prosódia e a entonação, presentes na fala, podem fornecer indicadores importantes sobre o fluxo de informação do discurso.

A transcrição é uma metodologia fundamental e importante para os futuros trabalhos do UPLA, pois permitirá que estudos sejam feitos com base em produção oral, possibilitando a análise de aspectos relevantes que possam futuramente auxiliar no ensino de PAL. Além disso, que venham a contribuir na elaboração de materiais didáticos adequados que apresentem a linguagem real do mundo acadêmico, permitindo que o professor elabore exercícios que aprimorem o vocabulário do aluno e sua produção oral.

Como argumenta Du Bois (1991, p.75) :

Através da experiência da transcrição, o transcritor está constantemente aprendendo sobre o discurso, não apenas adquirindo habilidade em discriminar as categorias implícitas no sistema de transcrição, mas também adquirindo uma imagem vívida da realidade conversacional que ele ou ela procura representar.<sup>50</sup>

A justificativa deste estudo reside exatamente no fato de que, tanto no próprio ato de transcrever quanto na transcrição resultante desse ato, o resultado é sempre positivo, pois a aprendizagem se torna bilateral. Não apenas os estudantes de PAL se beneficiam deste estudo, mas os próprios colaboradores e

---

<sup>50</sup>No original: Through the experience of transcribing the transcriber is constantly learning about discourse, not only gaining skill in discriminating the categories implicit in the transcription system but also acquiring a vivid image of the conversational reality that he or she is seeking to represent . (Tradução nossa).

responsáveis pela construção do corpus também aprendem sobre a própria língua.

Nesta seção, procurou-se definir o processo de transcrição e apresentar algumas de suas características e usos, e, ainda, apresentar a sua relevância para o presente estudo. Na próxima seção será apresentado o sistema de transcrição adotado neste trabalho.

#### 4.2.2 O Sistema de Transcrição de Du Bois

Uma conversa, uma reunião, uma palestra, uma aula, ou até mesmo algumas palavras trocadas em um bar podem se tornar objeto de análise de interação. Cada um desses eventos de fala apresenta um tipo de demanda, mas um bom sistema de transcrição deve ser capaz de acomodar os dados, com adaptação necessária se caso for, para a análise de interesse.

E como peça fundamental no processo de transcrição, está o transcritor. Du Bois (1991, p. 75) define o papel do transcritor apresentando a sua função no ato de transcrição e os ganhos por ele obtidos ao se envolver nesse processo:

Como há muito além do que o sistema de transcrição pode capturar, é o transcritor, imerso nas gravações dos eventos de fala e fundamentado na teoria do discurso, que está encarregado de fazer os ajustes necessários para fazer avançar o potencial do sistema de transcrição e o seu quadro teórico. [...] O processo de transcrição tem um potencial enorme de esclarecer os praticantes e de gerar neles um nível de percepção aguçado e um tipo de conhecimento íntimo que podem ser traduzidos em *insights* teóricos e em novos tipos de pesquisa.<sup>51</sup>

Com isso em mente, o sistema de transcrição deve contribuir para que o ato de transcrever seja, em si, uma experiência valiosa. E para que o tradutor possa manter a integridade dos dados e do contexto da fala, uma vasta gama de símbolos padronizados deve ser utilizada.

---

<sup>51</sup>No original: To the extent that there is more going on than the transcription system can capture, it is the transcriber immersed in the recorded speech event and grounded in discourse theory, who is in a position to rectify this, to advance the potential of the transcriptions system and its theoretical framework. [...] The process itself has tremendous potential for enlightening its practitioners, and for generating the level of keen perception and intimate knowledge that can translate into theoretical insights and new research directions. (Tradução nossa)

Os sistemas de transcrição, ou seja, os símbolos adotados para o registro da interação devem representar, ou pelo menos tentar disponibilizar para análise, as seguintes características (Du Bois, 1991, p. 76):

- palavras faladas, escritas de modo a permitir que cada item lexical seja reconhecido;
- a identificação do falante de cada turno;
- a sequência temporal dos enunciados, se estes seguem um ao outro em sucessão ou são simultâneos (como quando os falantes se sobrepõem);
- unidades básicas em que os enunciados foram articulados, tais como turnos e unidades de entonação;
- entonação de contorno, funcional ou foneticamente classificada;
- pronúncia;
- flutuações no tempo, tais como *tempo*, pausa e duração;
- ruídos não-verbais feitos por participantes do evento da fala, como um riso, um pigarro, uma inalação;
- qualidades especiais de voz que se estendem ao longo de um trecho de discurso;
- enunciados não-verbais que se tornam relevantes para a interação, tais como uma pessoa que serve comida aos participante, ou um súbito trovão ao fundo;
- metatranscrição e comentários da própria transcrição, indicando onde o transcritor está incerto das palavras faladas, e assim por diante;
- outras funções, conforme necessário.

No caso das transcrições deste estudo, que têm como objetivo registrar a fala-em-interação de aprendizes de PFA, pode ser necessário incluir traduções livres, uma vez que estamos nos reportando a falantes não nativos de português. Além disso, as transcrições podem apresentar dados contextuais gerais sobre os participantes do evento de fala, sobre a situação, a gravação do evento e a própria transcrição.

Com base no sistema de símbolos criado por Du Bois, é apresentada, a seguir, uma faixa de transcrições que exemplifica as características acima abordadas, e que, inclui, também, o sistema de símbolos adotados neste trabalho.

O exemplo mostrado é baseado no sistema presente em *Transcription Design Principles for Spoken Research* (Du Bois, 1991, p. 77).

**Figura 2 – Exemplo de Transcrição**

<p>(1)</p> <p>D: we have the used Dodge, but no new ones.</p> <p>G: ... No new ones? [Huh],</p> <p>D: [Dodge] is separate, from Chrysler Plymouth.</p> <p>G: ... You're [kidding me].</p> <p>D: [I mean] they're all affiliated, but th- th- [[they're not]] on our lot.</p> <p>G: [[Right]].</p> <p>...(.8) They're not on your lot. [Wow].</p> <p>D: [Yeah].</p>	<p>((CARSALES))</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------

Fonte: *Transcription Design Principles for Spoken Research* (Du Bois, 1991, p. 77)

No exemplo acima, a identificação do falante é apresentada através dos dois pontos posteriores à letra inicial do nome em maiúscula. As sobreposições na fala são indicadas através do uso de colchetes, com os colchetes da esquerda verticalmente alinhados. Dois casos separados de sobreposição em sucessão distinguem-se um do outro pelo uso de *um* versus *dois* colchetes, pausas de vários tipos são apresentadas através do uso de três pontos finais, com as durações das longas pausas apresentadas por um número entre parênteses. Palavras truncadas são apresentadas por um hífen no final da palavra. As unidades de entonação são marcadas por um *carriage return*; isso significa que cada linha contém uma unidade de entonação, e as entonações de contorno são amplamente classificadas, de acordo com as suas “*continuidades transicionais*”, por uma vírgula, ponto, ou ponto de interrogação no final da linha.

Os símbolos utilizados neste exemplo encontram-se em anexo. No próximo capítulo serão apresentadas amostras de transcrição com base no sistema de

símbolos adotado, bem como sugestões de análises que podem ser feitas a partir deste registro.

## 5. TRANSCRIÇÃO E PROPOSTAS DE ANÁLISE DE DADOS

### 5.1 A INTERFACE PRAGMÁTICA, AC E LdC.

A Linguística de Corpus nos permite verificar os padrões da língua, entretanto, não nos deixa chegar mais perto da interação e ver como ela realmente funciona. Para tanto, faz-se necessário estabelecer neste estudo uma interface entre a Linguística de Corpus, a Análise da Conversa e a Pragmática, pois a primeira é quantitativa e nos permitirá verificar o léxico nas combinações e na concordância; e as outras duas, qualitativas, permitindo o foco na interação, no contexto e na troca (Walsh, O’Keeffe, 2012).

A AC (Walsh, O’Keeffe, 2012) preocupa-se em analisar detalhadamente como a fala-em-interação é conduzida, em uma atividade por si só e como instrumento para o arranjo da ação social. Essa metodologia visa demonstrar que a conversa não é uma ação caótica e que as pessoas se organizam socialmente através da fala.

Como o propósito da AC é verificar de que forma os participantes colaboram na construção e na identificação de ações coerentes, devem ser gravadas as interações que ocorrem naturalmente, por isso os dados coletados deste tipo de interação são chamados de dados reais.

Tanto a AC quanto a LdC usam dados empíricos, ocorrências naturais da linguagem, levando em conta o contexto social. A análise do nível de tomadas de turno é crucial para a compreensão do contexto; é justamente nessas análises que podemos encontrar os insights mais reveladores. Entretanto, conforme comentado anteriormente, a LdC sozinha não nos permite investigar algumas características específicas do discurso oral, por isso a AC é tão importante, pois ela possibilita a análise detalhada, fornecendo descrições micro analíticas da fala-em-interação (Walsh, O’Keeffe, 2012).

Os autores Walsh e O’Keeffe melhor definem a AC como:

A abordagem CA baseia-se na noção de que os falantes estão constantemente conduzindo uma exibição social uns dos outros, e para os analistas, de seus estados cognitivos e de sua compreensão dos enunciados dos outros por meio de e por

referência à organização da sequência de turnos-de-fala e de reparação (2012, p.163)<sup>52</sup>

Já a Pragmática preocupa-se não apenas com os conjuntos de regras para sentenças bem formadas ou com significados inerentes dos signos, mas com a forma como a linguagem é usada em comunicação (Rühlemann, Aijmer, 2015).

A união da teoria Pragmática com as metodologias de LdC e AC nos permitirá analisar itens lexicais e contexto e verificar os padrões da linguagem acadêmica. Neste estudo, partiremos de uma proposta de transcrição que possibilitará a criação de um corpus oral de aprendizes de PLA e, posteriormente, a análise de fenômenos pragmáticos através da interação. Em outras palavras, com a elaboração do corpus, surge a possibilidade de verificar quais itens lexicais estão sendo utilizados com certa frequência, como estão sendo combinados e como estão sendo utilizados pelos falantes na estrutura conversacional.

A ideia é permitir que pesquisadores possam identificar, a partir do corpus, características linguísticas e interacionais que desempenham papel importante na construção dos enunciados e dos significados no meio acadêmico.

Embora a LdC permita uma análise muito mais rápida da frequência de palavras, das concordâncias e das palavras-chave do corpus, ela não consente ao pesquisador uma análise da dinâmica da interação (Morton, 2011), ou seja, fornece os dados, mas não os motivos pelos quais eles aparecem com certa frequência ou o porquê das combinações.

A tendência hoje é que os pesquisadores façam uso da LdC como uma ferramenta metodológica e agreguem a ela outras teorias para a análise da linguagem. Como afirma Morton (2011, p.03):

Quando o pesquisador grava, transcreve, anota e constrói um pequeno corpus falado contextualizado, um cenário diferente de possibilidades se abre em áreas que vão do léxico para áreas de

---

<sup>52</sup> No original: CA approach is based on the notion that interactants are constantly conducting a social display to each other, and to analysts, of their cognitive states and of their understanding of each others' utterances by means of and by reference to the organisation of turn-taking sequence and repair. (Tradução nossa).

uso (especialmente questões que envolvem a Pragmática, a Interação e o Discurso)<sup>53</sup>.

O foco de pesquisa do grupo UPLA reside, justamente, no uso que os aprendizes de PLA fazem da linguagem. Por isso, a combinação da Pragmática, da LdC e da AC é imprescindível, pois os pesquisadores do grupo não atentam apenas para a frequência das palavras e suas combinações, mas em como a conversa e a interação ocorrem no meio acadêmico e de que forma os itens lexicais, as expressões e as combinações aparecem quando a linguagem é utilizada neste ambiente - seja em debates, diálogos ou apresentações de trabalhos.

Walsh (2006, p. 51) afirma que:

Embora o foco original da CA seja em conversas que ocorrem naturalmente, é possível, em contextos institucionais específicos, onde as metas e ações dos participantes são claramente determinadas, que o valor das abordagens de AC possa ser vividamente mais percebido.<sup>54</sup>

A AC dá conta deste recado, usando as palavras de Walsh (2006, p. 52) para melhor exemplificar o uso desta metodologia na análise da interação em língua adicional em ambiente institucional:

A relevância de uma abordagem de AC para o contexto da sala de aula de L2 não é difícil de perceber. A AC tenta explicar as práticas no trabalho que permitem aos participantes em uma conversa dar sentido à interação e contribuir para ela. Possui paralelos claros: a conversa em sala de aula é composta por muitos participantes, e precisa apresentar transições suaves e expectativas claramente definidas para que os significados possam ser explícitos. Possivelmente, o papel mais significativo da AC é interpretar a partir dos dados e não impor categorias estruturais ou funcionais predeterminadas.<sup>55</sup>

---

<sup>53</sup>No original: When the researcher records, transcribes, annotates and builds a small contextualized spoken corpus, a different landscape of possibilities opens up in areas beyond lexis to areas of use (especially issues of pragmatics, interaction and discourse). (Tradução nossa).

<sup>54</sup>No original: Although the original focus of CA was naturally occurring conversation, it is perhaps in specific institutional settings, where the goals and actions of participants are clearly determined, that the value of CA approaches can be most vividly realized. (Tradução nossa).

<sup>55</sup>No original: The relevance of a CA approach to the L2 classroom context is not difficult to perceive. CA attempts to account for the practices at work that enable participants in a

A fala do autor representa mais uma das preocupações do grupo de pesquisa na análise das interações, é preciso “*interpretar a partir dos dados*”, ou seja, os dados coletados - a fala dos aprendizes em sala de aula – é que poderão nos dizer como a linguagem acontece no ambiente acadêmico, que tipo de léxico está sendo usado e como está sendo utilizado e, mais importante ainda, como o léxico e as estruturas da nova língua estão sendo utilizadas pelos participantes para que o significado e a organização conversacional sejam preservados e mantidos.

De acordo com Ochs (apud Rühlemann, Aijmer, 2015, p. 2):

O usuário do idioma escolhe uma forma lingüística variável de acordo com a situação social, que é amplamente concebida e inclui fatores como a identidade do falante, as relações com o ouvinte, o tipo de atividade e o posicionamento do falante.<sup>56</sup>

É pensando nesta coleta de dados, da interação em sala de aula que apresenta a estrutura conversacional e os elementos que a compõem que a combinação das metodologias da LdC e AC com a Teoria Pragmática se encaixam.

Corpora orais computadorizados podem fornecer informações sobre interlocutores e situações que permitam aos pesquisadores ver quem está dizendo o que a quem e em que contexto; a linguagem no contexto é o cerne da pesquisa pragmática (McAllister, 2015, p.29-30).<sup>57</sup>

Um corpus por si só, portanto, não serve apenas como armazenamento de dados. Ele é armazenado seguindo critérios que facilitam estudos não-

---

conversation to make sense of the interaction and contribute to it. There are clear parallels: classroom talk is made up of many participants, and there have to be smooth transitions and clearly defined expectations if meanings are to be made explicit. Possibly the most significant role of CA is to interpret from the data rather than impose predetermined structural or functional categories. (Tradução nossa).

<sup>56</sup> No original: The language user chooses a linguistic form variably according to the social situation, which is broadly conceived and includes such factors as speaker identity, relations to the hearer, activity type and speaker stance. (Tradução nossa).

<sup>57</sup> No original: [...] computerized spoken corpora can provide information on interlocutors and situations allowing researchers to see who is saying what to whom, and in what context; and language in context is at the heart of pragmatics research. (Tradução nossa).

lineares, tanto de forma qualitativa quanto de forma quantitativa, pois quando acessado através de software específico para o estudo da linguagem, permite fazer vários tipos de observação.

Disponibilizar estes dados através da transcrição significa auxiliar no estabelecimento e desenvolvimento de uma área de pesquisa dedicada ao PFA, com materiais didáticos específicos para este fim, que auxiliem na produção acadêmica do estudante de português e que facilitem a sua aprendizagem.

Na próxima seção serão apresentadas as amostras de transcrição e as possibilidades de análises realizadas através da AC com a Teoria Pragmática com base nos dados coletados a partir da interação da sala de aula de português como língua adicional.

## 5.2 TRANSCRIÇÃO E PROPOSTAS DE ANÁLISES

As passagens selecionadas e apresentadas nesta seção compõem uma pequena amostra do tipo de transcrição adotada, dos símbolos utilizados e de uma possível análise com base na perspectiva teórico-metodológica da AC e da Teoria Pragmática.

Os trechos apresentados foram escolhidos tomando como base a teoria Pragmática, considerando partes interessantes para a análise do uso da linguagem em ambiente acadêmico geral. Portanto, o que se pretende, a partir dos trechos, é mostrar como foi realizada a transcrição, o tipo de interação que estava acontecendo e como a linguagem é utilizada pelos aprendizes. Tudo isso com o intuito de estabelecer aqui um protótipo de análise, de sugerir futuras pesquisas na área e apresentar possibilidades de análises que podem ser feitas a partir da transcrição.

Na primeira amostra, apresenta-se a transcrição de um debate na sala de aula de PFA acerca dos benefícios e malefícios de um chá conhecido como *Ayahuasca*. A transcrição apresentada representa a interação dos estudantes e do professor tutor na discussão do assunto em questão. Inserida em uma tabela que garante uma melhor organização do material, a transcrição revela apenas um trecho da interação, mas possui detalhes interessantes de serem analisados sob a perspectiva da Teoria Pragmática.

Para garantir uma melhor organização do trabalho, são apresentados o contexto dos dados e os participantes da interação. Os dados gerados aparecem em forma de tabela conforme a transcrição adotada. O uso de símbolos e a sua representatividade são abordados após a tabela de dados, seguidos pelas sugestões de análises com base na Pragmática em conjunto com a AC.

### Amostra 1

**Contexto:** Em uma das aulas de PFA da PUCRS, a professora propõe aos alunos um debate sobre um tema polêmico na época, o *Ayahuasca*. Nesta aula, os alunos discutem os benefícios e os malefícios do chá.

**Participantes:** Participam do debate o tutor e cinco alunos de uma turma de PFA, composta por um venezuelano, três peruanos e um nepalês. Os nomes dos participantes são preservados, identificados por letras maiúsculas aleatórias do alfabeto. O professor tutor é indicado pela letra (**T**).

**Figura 3** – Transcrição de um debate em uma aula de PLA.

(1) (( PPE))

**L:** Não sei exatamente como se chama aqui,

**M:** Xamã.

**T:**Xa- Xamã.

**L:** Xamã.

**T:**Uhum.

**L:**E tem que ser uma..

..é..preparação por alguém que conhece,

e..o mesmo o mesmo ritual,

tem que ser guiado por uma pessoa também que conhece,

que é um Xamã,

só para adicionar isso aí né,

..que é compartilhado tanto aqui como no Peru.

**T:**Uhum.

**M:** é,

..e adicionalmente <X X>,

assim..na Venezuela também serve <X X>,

da Amazônia,

[Amazônia do Brasil].

**G:** [Colômbia]

Fonte: a autora

Na transcrição apresentada acima, o nome dos participantes é preservado e substituído por letras que aparecem em maiúsculo. A disposição da identificação do falante precedendo o símbolo de dois pontos(:) serve para diferenciar o produtor da fala do próprio enunciado.

Os enunciados, de acordo com o sistema de transcrição adotado, devem ser apresentados por unidades entoacionais, ou seja, um filete de discurso

proferido no âmbito de um único contorno entoacional coerente. O uso do ponto final (.) nos enunciados mostra que o discurso proferido foi finalizado. Já o uso de vírgula (,) ao final do enunciado marca a continuidade entre um enunciado e o próximo. As pausas na fala são indicadas através do uso de pontos, três pontos (...) são utilizados para indicar que a pausa é média, dois pontos (..) são utilizados para indicar uma pausa curta, de 0,2 segundos aproximadamente. No exemplo acima, há o registro de algumas pausas curtas entre as falas dos participantes.

O uso do hífen seguido de palavra não finalizada (na passagem transcrita, a palavra Xamã aparece transcrita como Xa-) mostra que a palavra foi truncada, ou seja, o falante abandona a palavra e dá início a uma nova tentativa de produção. O uso do símbolo (<X X>) serve para mostrar que o transcritor não conseguiu compreender o que foi dito, ou seja, não está certo da palavra utilizada pelo falante. Para marcar a sobreposição na fala, foram utilizados colchetes que aparecem sobrepostos e alinhados à esquerda do enunciado do falante.

De uma perspectiva interacional, alguns elementos podem ser destacados e, posteriormente, profundamente analisados, como a questão dos turnos de fala, por exemplo, ou também da própria organização interacional, pois se percebe que, nos dados apresentados acima, tanto os alunos quanto o professor colaborativamente negociam significados e mutuamente constroem uma compreensão acerca do que está sendo discutido. A organização interacional estrutura-se por uma fala mais longa produzida por um dos estudantes (L), uma vez que se propõe a explicar como funciona a preparação do chá *Ayahuasca*. Enquanto o falante (L) explica sobre o chá, os outros participantes da interação produzem respostas curtas, por vezes auxiliando na produção da fala, principalmente para confirmar o vocabulário utilizado.

O professor executa, no excerto acima, não o papel de questionador, mas o de motivador, confirmando o que está sendo falado, como uma forma de garantir ao estudante que sua produção está correta. Entende-se, também, que o posicionamento do professor não se dá exclusivamente à compreensão do significado, mas do vocabulário que está sendo empregado no português, em outras palavras, o professor confirma que a produção do aprendiz em língua portuguesa está sendo correta, efetiva e garantindo a compreensão.

Percebe-se que o professor (**T**) faz a confirmação do que está sendo dito pelo uso repetido da expressão *uhum*, elemento que poderia ser estudado a partir da Teoria dos Atos de Fala, a qual assume que os enunciados podem ser descritos de acordo com o as ações que eles desempenham (O’Keeffe et al., 2011)

Segundo O’Keeffe:

A teoria dos Atos de Fala fornece uma taxonomia das diferentes funções que os enunciados podem executar, e oferece, também, uma aproximação para compreender a discrepância aparente entre o que nós dizemos e o que nós queremos significar (2011, p.84).<sup>58</sup>

A partir dos dados gerados, pode-se, então, analisar a fala através do contexto, ou seja, o que os alunos querem significar ao utilizarem certos elementos linguísticos e se esses elementos utilizados estão sendo empregados da maneira correta.

Estudos sobre competência pragmática também são realizados a partir da análise com base em corpora, principalmente em contextos translinguísticos (Kasper e Rose, 2002). Nesse sentido, percebe-se que no trecho transcrito os alunos estão conduzindo o diálogo. Há trocas de informação por parte dos integrantes sem que haja interrupções para retomar enunciados mal compreendidos, ou seja, os alunos demonstram que possuem competência pragmática e interacional para desenvolver e falar sobre o tópico em questão.

Assim, o trecho transcrito abre portas para o estudo dessas competências, exigidas em exames de proficiência, como o Celpe-Bras, “um exame que, alinhando-se a perspectivas teóricas que compreendem a linguagem como ação, busca avaliar a proficiência em português a partir do uso da linguagem como prática social (Niederauer, 2014, p.406)”

A Competência Interacional (CI) é uma abordagem que, segundo Niederauer (2014, p.408) se interessa em:

---

<sup>58</sup> No original: Speech Act Theory provides a taxonomy of the different functions that utterances might perform, and it also offers an approach to understanding the apparent discrepancy between what we say and what we mean. (Tradução nossa).

como as práticas sociais e discursivas se dão na interação e em como o significado é estabelecido, negociado, mantido e transformado na interação. [...] Analisar a interação ou avaliar essa competência implica, assim, considerar o que está sendo realizado pelos falantes em determinado contexto, por meio de práticas sociais e discursivas [...] as habilidades orais não existem dentro do cérebro do falante bilíngue, tampouco pertencem ao indivíduo; antes disso, elas são construídas de modo conjunto pelos falantes engajados na interação.

O foco dessa abordagem, portanto, não está no indivíduo, mas nos falantes engajados na interação. Ou seja, é importante desenvolver no aprendiz esta competência, que poderá ser aprimorada quando mais materiais forem criados para auxiliar o aluno que quer aprender PLA. Do contrário, sem materiais e pesquisas que possam auxiliar nas análises das dificuldades enfrentadas por estes aprendizes, o ensino será precário, focado na gramática e não no uso real da língua. Por isso, reforça-se aqui a importância de trabalhos desenvolvidos a partir de corpus, pois eles representam e disponibilizam os usos reais da língua.

## Amostra 2

**Contexto:** Ao ingressarem na universidade, os alunos podem ser requisitados a elaborar um artigo como trabalho final para a disciplina que estão cursando, ou até mesmo um Trabalho de Conclusão. Assim, os alunos fazem uma orientação com os professores acerca desses trabalhos a serem desenvolvidos. O trecho a seguir apresenta a transcrição de um desses encontros de orientação em que são discutidos os aspectos sobre o trabalho e apontadas algumas considerações.

**Participantes:** Participam da orientação, o professor o tutor e uma das alunas do curso de PPE, de nacionalidade chinesa. O nome da aluna é preservado, identificado por uma letra maiúscula aleatória do alfabeto. O professor tutor é indicado pela letra (T).

**Figura 4 – Orientação de Trabalho**

(2) ((PPE))

**A:** eu quero começar com uma..piadinha,  
tipo..é minha prima Guarani,  
mas é eu né?,  
sou eu.

**T:** Hum.

**A:** É possível?,  
Vai para uma universidade Chinesa.

**T:** Uhm.

**A:** Desculpa..é..  
.. uma..ah..sim

**T:** Para uma universidade Chinesa é possível?

**A:** Mas é [ muito ],

**T:** [ ir ],

**A:** [ Ir ].

**T:** [ Ir ],  
Porque tu tá,  
Tu usou um gerúndio aí né?,  
É possível ir para uma universidade Chinesa?

Fonte: a autora

Na transcrição apresentada acima, o nome dos participantes é preservado e substituído por letras que aparecem em maiúsculo. A disposição da identificação do falante precedendo o símbolo de dois pontos (:) serve para diferenciar o produtor da fala do próprio enunciado.

Os enunciados são mantidos de com as unidades entoacionais, um filete de discurso proferido no âmbito de um único contorno entoacional coerente. O uso do ponto final (.) nos enunciados mostra que o discurso proferido foi finalizado. Já o uso de vírgula (,) ao final do enunciado marca a continuidade entre um enunciado e o próximo. As pausas na fala são indicadas através do uso de pontos, três pontos (...) são utilizados para indicar que a pausa é média, dois pontos (..) são utilizados para indicar uma pausa curta, de 0,2 segundos aproximadamente. No exemplo acima, há o registro de algumas pausas curtas entre as falas dos participantes.

Para marcar a sobreposição na fala, devem ser usados colchetes alinhados à esquerda e sobrepostos, para indicar que o enunciado foi proferido ao mesmo tempo.

Do ponto de vista interacional, a estudante inicia a fala através de uma explicação, mostrando ao professor orientador o que ela tentou significar com o uso da palavra “Guarani”. A aluna utiliza o termo “piadinha” para fazer referência a uma foto presente nos slides do trabalho que ela está desenvolvendo. Na foto, a aluna aparece montada em um cavalo, por isso ela explica o termo ao professor orientador. O uso do termo “piadinha” refere-se a essa “brincadeira” com as identidades, já que a aluna é de nacionalidade chinesa.

**Figura 5 – Fotografia**



Fonte: Conferência Internacional para a Melhoria Contínua da Qualidade do Ensino de Português como Língua Estrangeira/2016

Ambos os participantes colaboram na construção da interação, mantendo o diálogo e garantindo a competência pragmática, melhor explicada nas palavras de O’Keefe et al. (2011, p.138):

A competência pragmática envolve competências pragmlinguísticas e sociopragmáticas. Pragmlinguística refere-se ao conhecimento da gama de opções disponíveis para a realização de várias ações pragmáticas e a sociopragmática refere-se ao conhecimento de como selecionar uma escolha adequada dada uma meta particular em um cenário particular.<sup>59</sup>

A competência pragmática pode ser, também, futuramente explorada a partir de trechos transcritos como o apresentado acima. Ela está relacionada ao uso que o falante faz da língua, de forma apropriada ao ambiente em que está inserido, onde contam os aspectos sócio-culturais. Se o uso que a estudante faz da linguagem garante a comunicação e a interação entre os participantes, ou seja, se não há pausas na comunicação para retomada de enunciados mal compreendidos ou para retomada de emprego confuso de vocabulário, então o diálogo é mantido, mostrando que a aluna possui competência pragmática em sua comunicação.

Um outro aspecto interessante para futura investigação é o uso de partículas expressivas, ou marcadores discursivos, como é o caso de “tipo”, que aparecem em narrativas orais, mostrando as funções discursivo-interacionais que exercem em determinados contextos. Um estudo sobre as funções destas partículas (Castelano e Ladeira, 2010) apresentou três diferentes finalidades marcantes analisadas a partir da AC: marcar sequenciamento narrativo, inserir sequências explicativas e marcar hesitação.

As autoras da pesquisa concluíram que a utilização dessas expressões está associada à busca de um grau maior de compreensão entre os interlocutores de um discurso oral. Seria interessante, portanto, analisar esses marcadores em contextos em que a produção oral é proveniente de ambiente acadêmico, para verificar se eles teriam as mesmas finalidades e, também, se, mesmo em ambiente caracterizado por fala mais formal, aparecem com tanta frequência.

Quanto à função dos marcadores discursivos, Fávero (2005, p.47) acrescenta que:

---

<sup>59</sup> No original: “Pragmatic competence involves both pragmlinguistic and sociopragmatic competences. Pragmlinguistics refers to knowledge of the range of options available for performing various pragmatic actions and sociopragmatics refers to the knowledge of how to select an appropriate choice given a particular goal in a particular setting.” (Tradução Nossa)

Considerando as funções e argumentativas dos marcadores conversacionais, verificamos que esses elementos desempenham papel de especificadores, coordenadores, subordinadores, entre outros. Por meio desses marcadores, podem-se explicar os deslocamentos referenciais locais ou globais com a função de conduzir e orientar as atividades do locutor e do seu interlocutor.

A autora afirma ainda que esses marcadores discursivos têm como função sustentar o tópico e auxiliar os interlocutores a darem continuidade ao assunto, aspectos interessantes de serem investigados com base em corpora.

Apesar de parecerem pequenos, os trechos transcritos nesta pesquisa apresentam elementos linguísticos variados que podem ser futuramente explorados e profundamente analisados. Além de podermos verificar a interação entre os participantes, analisada com base na metodologia da AC, de analisar marcadores discursivos, apresentados na transcrição pelo termo “tipo”, e estudar elementos linguísticos através dos Atos de Fala, pode-se ainda, explorar a frequência com que certos termos aparecem nos dados gerados ou as escolhas linguísticas dos falantes para a comunicação. Por isso estudos com foco na Pragmática são tão relevantes, como argumentam Kasper e Rose (2002, p. 2-3):

Os significados pragmáticos surgem de escolhas entre formas linguísticas, como usar um marcador discursivo ou uma partícula sobre a outra, ou optar por um formato linguístico de um ato comunicativo em vez de uma alternativa contextualmente possível para transmitir força ilocucionária ou polidez. As escolhas dos falantes e dos escritores, entretanto, não são irrestritas, mas são governadas por convenções sociais, que podem ser flexionadas em graus diferentes, variando contextualmente, mas apenas deixam totalmente de lado o perigo de perder reivindicações para , propriedade interna ou sanidade.<sup>60</sup>

Através da transcrição da fala em interação de aprendizes de português como LA em ambiente acadêmico, elementos linguísticos, dos mais variados

---

<sup>60</sup> No original: Pragmatic meanings arise from choices between linguistic forms, such as using one discourse marker or particle over another, or opting for one linguistic format of a communicative act instead of a contextually possible alternative to convey illocutionary force or politeness. Speakers' and writers' choices, however, are not unconstrained but are governed by social conventions, which can be flexed to different, contextually varying degrees but only entirely set aside the peril of losing claims to face, insider status, or sanity. (Nossa Tradução).

tipos, poderão ser analisados e estudados. Com a percepção do tipo de linguagem que está sendo utilizada pelo aluno estrangeiro poderemos aprimorar o próprio ensino da pragmática.

O ensino da pragmática na sala de aula é de suma importância. Primeiro porque existe a necessidade desse tipo de ensino para garantir a adequação ao gênero acadêmico em questão; segundo, porque esse ensino é efetivo. No mesmo sentido, o ensino da pragmática contribui para prevenir a falha pragmática, pois por mais que o aluno seja exposto à LA, ele não adquire a competência pragmática, pois essa exposição apenas é insuficiente. (O’Keeffe et al., 2011).

Assim, as amostras aqui apresentadas têm por objetivo despertar o pesquisador para estes aspectos da língua, levando em consideração que o estudante deve aprender a usar a língua de maneira adequada, tendo consciência do emprego das palavras no sentido pragmático, tomando o contexto, também, como base para a produção linguística. No próximo capítulo, serão apresentadas as considerações finais dos elementos apresentados nesta pesquisa.

## 6 CONSIDERAÇÕES FINAIS

O presente estudo surgiu através da observação do aumento de estrangeiros em mobilidade acadêmica que chegam à universidade para estudar o português. Visto que os estudos de PLA vinham assumindo uma posição de destaque no cenário atual, esse trabalho quis contribuir com os estudos do projeto UPLA (Uso e Processamento de Língua Adicional) que tem como objetivo auxiliar o desenvolvimento da competência linguística em PFA desses estudantes.

No primeiro capítulo, foram apresentados e discutidos os conceitos e as definições de Língua Adicional e suas características, assim como a necessidade de criação de uma área que atenda os alunos na produção escrita e oral e para fins específicos, como exige o meio acadêmico. No segundo capítulo desta pesquisa, dedicou-se à apresentação da Teoria Pragmática, argumentando a importância desta perspectiva no desenvolvimento de estudos nesse âmbito. Além disso, foram apresentados os conceitos da Análise da Conversa, perspectiva teórico-metodológica que analisa a fala-em-interação.

O terceiro capítulo dedicou-se a explorar conceitos e definições da Linguística de Corpus, mostrando, também, os tipos de corpora existentes, justificando a escolha pelo corpus a ser elaborado, o corpus de aprendiz. O quarto capítulo deste trabalho, apresentou a ideia geral do que corpus que se pretende montar e o tipo de transcrição adotada, com suas características e finalidades.

O quinto capítulo deste trabalho apresentou a transcrição de alguns trechos da interação de aprendizes de PLA com sugestões para futuras análises de elementos linguísticos e interacionais que ali apareciam.

Como forma de contribuir para o projeto UPLA, o presente trabalho apresentou uma proposta de transcrição para a interação da sala de aula de português como LA, dentro do gênero acadêmico, ou seja, PFA para alunos de mestrado e doutorado que necessitam produzir textos orais e escritos em português. A transcrição adotada parte do princípio de que o transcritor é peça fundamental na interpretação dos dados registrados, e tem como principal objetivo permitir aos pesquisadores a análise de elementos linguísticos, da competência pragmática, da interação, do estudo do contexto, dos turnos de fala, entre outras informações e características. Em outras palavras, esse tipo de

transcrição é ideal para os estudos da Análise da Conversa, que tem como foco estudar como os falantes se organizam em uma comunicação, partindo da metodologia para a análise de padrões linguísticos explorados através da teoria Pragmática.

Quando colocamos a AC em uma interface com a Linguística de Corpus (LC) possibilitamos a análise da interação e das expressões que são mais utilizadas no estabelecimento da organização da fala na linguagem acadêmica. Ao oferecer esse arcabouço teórico, a presente pesquisa trará subsídios para os demais trabalhos do grupo, uma vez que o UPLA iniciou, a partir dela, seus estudos na interface da oralidade.

Tendo definido a metodologia de transcrição, abre-se a porta para os estudos da interação em sala de aula de português como LA. Isso permitirá não apenas uma análise qualitativa dos dados registrados, mas também, uma avaliação de como a língua portuguesa é utilizada pelos estrangeiros, que tipos de construções eles produzem, o tipo de vocabulário e, principalmente, onde residem suas dificuldades, para que, assim, materiais didáticos possam ser elaborados atendendo às necessidades reais desses alunos.

E quando falamos sobre materiais que atendam às necessidades desses aprendizes, estamos falando de língua em uso, ou seja, materiais que capacitem o aluno à competência na linguagem acadêmica, ou seja, que permitam que esse aluno adquira, produza, e que saiba utilizar essa linguagem ao se posicionar na sala de aula, ao escrever um artigo, ao elaborar e apresentar uma palestra.

Espera-se que este trabalho contribua para os estudos referentes ao discurso acadêmico do português brasileiro, mais especificamente no que se refere à investigação de padrões linguísticos usados por aprendizes deste gênero. Também ressalta-se que este trabalho é incipiente, mas de grande valia, e o que foi apresentado até o momento é um protótipo de pesquisa, uma parte do todo que almejamos enquanto pesquisadores da área.

Os objetivos e ideias apresentados nesta pesquisa serão posteriormente mais aprofundados. Estabelecida a transcrição para registro dos dados, a ideia é seguir os passos para a construção do acervo linguístico do português como LA para, depois, disponibilizar os dados para futuras pesquisas. Estudos esses que podem partir das sugestões de análises apresentadas nesta pesquisa.

Uma aplicação futura das sugestões de análise propostas nesta pesquisa poderá colaborar para ensino e aprendizagem de alunos brasileiros e estrangeiros que buscam se aperfeiçoar na língua portuguesa, contribuindo para a crescente área de Português para Fins Acadêmicos.

## REFERÊNCIAS

AIJMER, K.; RÜHLEMANN, C. *Corpus Linguistics: A Handbook*. Cambridge: Cambridge University Press, 2015.

BARDOVI-HARLIG, K.; DÖRNYEI, Z. Do Language Learners Recognize Pragmatic Violations? Pragmatic Versus Grammatical Awareness in Instructed L2 Learning. *Tesol Quarterly* 32.2 , p. 233-259, 1998.

BARDOVI-HARLIG, K.; MAHAN-TAYLOR, R. *Teaching Pragmatics*. Washington DC: US: Department of State Office of English Language Programs, 2003

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press, 1996.

BIBER, D.; SUSAN, C. *Register, Genre and Style*. Cambridge: Cambridge University Press, 2009.

CASTELANO, K.L.; LADEIRA, W.T. Funções discursivo-interacionais das expressões “assim”, “tipo” e “tipo assim” em narrativas orais. *Em: Revista de Divulgação Científica em Língua Portuguesa, Linguística e Literatura*. Ano 06, n.12, 2010.

CELCE-MURCIA, M.; OLSHTAIN, E. *Discourse and Context in Language Teaching: A Guide for Language Teachers*. Cambridge: Cambridge University Press, 2000.

CLIFT, R.; DREW, P.; HUTCHBY, I. Conversation Analysis. In: Östman, Jan-Ola, and JefVerschueren, eds. *The pragmatics of interaction*. Vol. 4. John Benjamins Publishing, p. 40-50, 2009.

CRANDALL, E.; BASTURKMEN, H. Evaluating pragmatics-focused materials. *ELTJournal*, 58(1): p.38–49, 2004.

CUMMINS, J. Putting language proficiency in its place: Responding to critiques of the conversational/academic language distinction. In: J. Cenoz & U. Jessner (Eds.) *English in Europe: The acquisition of a third language*. Clevedon: Multilingual Matters, 2000

DEPPERMAN, A.; SCHÜTTE, W. Data and transcription (Ch. 8). In: *Handbook of Interpersonal Communication* , p.179, 1998.

DU BOIS, J. W. Transcription design principles for spoken discourse research. *Pragmatics* 1.1:p. 71-106, 1991

DU BOIS, J. W.; SCHUETZE-COBURN, S.; PAOLINO, D.; CUMMING, S. Outline of discourse transcription. In Edwards, Jane A., and Martin D. Lampert. *Talking data: Transcription and coding in discourse research*. Psychology Press, p.45-90, 2014.

FÁVERO, L. L.; ANDRADE, M. L. C. V. O.; AQUINO, Z. G. O. *Oralidade e escrita: perspectivas para o ensino de língua materna*. 5 ed. São Paulo: Cortez, 2005.

GRANGER, S.; HUNG, J.; PETCH-TYSON, S.; eds. *Computer learner corpora, second language acquisition, and foreign language teaching*. Vol. 6. John Benjamins Publishing, 2002.

HUANG, Y. *Pragmatics*. Oxford: Oxford University Press, 2007.

HUNSTON, S. *Corpus linguistics*. *Linguistics* 7.2: p.215-244, 2006.

HUTCHBY, I.; WOOFFITT, R. *Conversation analysis: principles, practices and applications*. Cambridge: Polity Press, 1998.

JORDAN, R. R. *English for academic purposes: a guide and resource book for teachers*. Cambridge: Cambridge University Press, 1997.

JUDD, E. L.; TAN, L.; WALBERG, J. H. *Teaching additional languages*. 2001.

KASPER, G; ROSE, K. R. *Pragmatic Development in a Second Language*. Oxford: Blackwell Publishing, 2002.

KASPER, G; WAGNER, J. *Conversation Analysis in Applied Linguistics*. *Annual Review of Applied Linguistics*, 34: p. 1-42. Cambridge University Press, 2014

KENNEDY, G. *An introduction to Corpus Linguistics*. New York: Addison Wesley Longman Limited, 1998.

KERBRAT-ORECCHIONI, C. *Análise da Conversação: princípios e métodos*. São Paulo: Parábola Editorial, 2006.

LEECH, G. *Principles of Pragmatics*. Longman: 1983

LEVINSON, S. C. *Pragmática*. São Paulo: WMF Martins Fontes, 2007.

MCALLISTER, P.G. *Speech Acts: a synchronic perspective*. In: AIJMER, K.; RÜHLEMANN, C. *Corpus Linguistics: A Handbook*. Cambridge: Cambridge University Press, p. 29-51, 2015.

MEUNIER, F. *The pedagogical value of native and learner corpora in EFL grammar teaching*. In: GRANGER, S.; JOSEPH, H.; PETCH-TYSON, S. eds. *Computer learner corpora, second language acquisition, and foreign language teaching*. Vol. 6. John Benjamins Publishing, p 119-142, 2002.

MOLSING, K. V.; PERNA, C. *Research and Teaching in Portuguese for Specific Purposes*. *BELT-Brazilian English Language Teaching Journal* 5.2: p.1-7, 2007.

MORO, B. I. **Advérbios de posicionamento em textos escritos de Português**

**Academico.** 2014. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014.

MORTON, T. Analysing university spoken interaction: a CL/CA approach. *International journal of corpus linguistics* 16.3, p. 325-345, 2011.

NIEDERAUER, M. *Competência Interacional: critério para avaliação da produção oral em língua adicional*. Universidade de Brasília. Brasília, 2014.

OCHS, E. Linguistic resources for socializing humanity, 1996 *In: AIJMER, K.; RÜHLEMANN, C. Corpus Linguistics: A Handbook*. Cambridge: Cambridge University Press, p.2, 2015.

O'KEEFE, A.; MCCARTHY, M.; CARTER, R. *From corpus to classroom*. Cambridge: Cambridge University Press, 2007.

O'KEEFFE, A.; CLANCY, B.; SVENJA, A. *Introducing pragmatics in use*. Taylor & Francis, 2011.

**PACHECO, A. A Aquisição de Morfemas em Inglês como L2: Uma análise dos padrões evolutivos através do BELC (Brazilian English Learner Corpus)**. 2010. Tese (Doutorado em Teoria e Análise Linguística) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

PIMENTEL, C.L. A tradução da palavra saudade: uma pesquisa baseada em corpus. Monografia (Finalização do Curso de Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2008

PINTO, J. P. Pragmática. Introdução à linguística: domínios e fronteiras, volume 2/ Fernanda Mussalim, Anna Cristina Bentes, organizadoras. 8ª edição. São Paulo: Cortez, 2012

**ROJAS, J. P. Processo de fossilização na interlíngua de hispanofalantes aprendizes de português no Brasil: acomodação consentida?**. 2006. Dissertação (Mestrado em Linguística Aplicada)-Universidade de Brasília, Brasília, 2006.

SANTOS, P. O ensino de português como segunda língua para falantes de espanhol: teoria e prática. Em CUNHA, MJ e SANTOS, P.(orgs.) *Ensino e Pesquisa em Português para Estrangeiros*. Editora UnB, Brasília, Brazil, 1999.

**SANTOS, S.N. Uma Análise dos Substantivos como Marcadores de Posicionamento em Artigos Acadêmicos em Língua Portuguesa**. 2015. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Católica do Rio Grande do Sul, Porto Alegre, 2015.

SARDINHA, T. B. *Linguística de Corpus*. Barueri, São Paulo: Manole: 2004.

SCHLATTER, M, GARCEZ, P. *Línguas adicionais na escola: aprendizagens colaborativas em inglês*. Edelbra Editora Ltda, 2012.

STERN, H. H. *Fundamental Concepts of Language Teaching*. Oxford University Press, 1991.

TORRES, L. S. **Escrita Científica em Português para Hispano Falantes: recursos linguísticos-computacionais baseados em métodos de alinhamento de textos paralelos**. 2015. Tese (Doutora em Ciências da Computação e Matemática Computacional) – Instituto de Ciências de Matemática e de Computação, Universidade de São Paulo, São Carlos, 2015.

TORRES, L.S.; ALUISIO, S.M.; DURAN, M.S. *O Tratamento de Marcadores Discursivos em uma Ferramenta de Apoio à Escrita Acadêmica em Português para Nativos de Espanhol*. Em: Conference: Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish (PROPOR 2014), At São Carlos, Brasil. Paper...São Carlos, 2014. Disponível em: <<https://www.researchgate.net/publication/282504772> O Tratamento de Marcadores Discursivos em uma Ferramenta de Apoio a Escrita Acadêmica em Português para Nativos de Espanhol> Acesso em: 09 set.

VELLENGA, H. *Learning pragmatics from ESL and EFL textbooks: How likely?*, TESLJ8(2), 2004.

WALSH, S. *Investigating Classroom Discourse*. New York: Routledge, 2006

WALSH, S. O'KEEFFE, A. Applying corpus linguistics and conversation analysis in the investigation of small group teaching in higher education. *Corpus Linguistics and Linguistic Theory*, v. 8, n. 1, p. 159-181, 2012.

WALSH, S. *Classroom discourse and teacher development*. Edinburgh University Press, 2013.

WOOFFITT, R. *Conversation Analysis and Discourse Analysis: A Comparative and Critical Introduction*. SAGE Publications, 2005

YULE, G. *Pragmatics*. Oxford University Press, 1996.

## **ANEXOS**

## ANEXO 1 – SISTEMA DE SÍMBOLOS DE DU BOIS

UNITS	
Intonation unit	<i>{carriage return}</i>
Truncated intonation unit	--
Word	<i>{space}</i>
Truncated word	-
SPEAKERS	
Speaker identity/turn start	:
Speech overlap	[ ]
TRANSITIONAL CONTINUITY	
Final	.
Continuing	,
Appeal	?
TERMINAL PITCH DIRECTION	
Fall	\
Rise	/
Level	-
ACCENT AND LENGTHENING	
Primary accent	^
Secondary accent	‘
Booster	!
Lengthening	=
TONE	
Fall	\
Rise	/
Fall-rise	\ /
Rise-fall	/ \
Level	-
PAUSE	
Long	...(N)
Medium	...
Short	..
Latching	(0)
VOCAL NOISES	
Vocal noises	( )
Inhalation	(H)
Exhalation	(Hx)
Glottal stop	%
Laughter	@

QUALITY	
Quality	<Y Y>
Laugh quality	<@ @>
Quotation quality	<Q Q>
Multiple quality features	<Y<Z Z>Y>
PHONETICS	
Phonetic/phonemic transcription	(/ /)
TRANSCRIBER'S PERSPECTIVE	
Researcher's comment	(( ))
Uncertain hearing	<X X>
Indecipherable syllable	X
SPECIALIZED NOTATIONS	
Duration	(N)
Intonation unit continued	&
Intonation subunit boundary	
Embedded intonation unit	<   >
Reset	{ <i>Capital Initial</i> }
False start	< >
Codeswitching	<L2 L2>
NON-TRANSCRIPTION LINES	
Non-transcription line	\$
Interlinear gloss line	\$G
RESERVED SYMBOLS	
Phonemic/orthographic	'
Morphosyntactic coding	+ * # { }
User-definable	" ~ ;