

RESSALVA

Atendendo solicitação da autora,
o texto completo desta tese será
disponibilizado somente a partir
de 11/03/2021



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Queila Barbosa Lopes

MulTeC:

A construção de um *corpus* multimodal em teletandem

São José do Rio Preto
2019

Queila Barbosa Lopes

MulTeC:

A construção de um *corpus* multimodal em teletandem

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientadora: Prof^a. Dr^a. Solange Aranha

São José do Rio Preto
2019

L864m Lopes, Queila Barbosa
MulTeC : A construção de um corpus multimodal em
teletandem / Queila Barbosa Lopes. -- São José do Rio
Preto, 2019
161 p. : il., tabs.

Tese (doutorado) - Universidade Estadual Paulista
(Unesp), Instituto de Biociências Letras e Ciências Exatas,
São José do Rio Preto
Orientadora: Solange Aranha

1. Linguística Aplicada. 2. Ensino à distância. 3.
Linguagem e línguas Estudo e ensino. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do
Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados
fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Queila Barbosa Lopes

MulTeC:

A construção de um corpus multimodal em teletandem

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadoras: CAPES

Comissão Examinadora

Prof^a. Dr^a. Solange Aranha
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Antônio Paulo Berber Sardinha
PUC – São Paulo

Prof. Dr. Júlio César Rosa de Araújo
UFC – Fortaleza

Prof^a. Dr^a. Claudia Zavaglia
UNESP – Câmpus de São José do Rio Preto

Prof^a. Dr^a. Suzi Spatti Cavalari
UNESP – Câmpus de São José do Rio Preto

São José do Rio Preto
11 de março de 2019

À Solange Aranha que confiou-me o desafio de organizar um *corpus*
com os dados do Banco TTDii.
Minha eterna gratidão!

AGRADECIMENTOS

A Deus, minha força e fortaleza, sem as quais a depressão teria me impedido de concluir essa etapa da minha vida com algum êxito.

À UFAC, por meu afastamento com ônus para realização de meus estudos de doutoramento, sem o qual esta pesquisa não teria sido realizada.

À minha orientadora, Prof.^a Dr.^a Solange Aranha, pelas inúmeras leituras cautelosas e sempre bastante criteriosas de meus textos, bem como pelos momentos em que ouviu minhas dificuldades para além de pesquisadora, como aquelas que enfrentei no papel de esposa e mãe.

À Prof.^a Dr.^a Ciara Wigham, cuja bondade, competência e generosidade não podem ser descritas em palavras. Agradeço por tantas vezes, mesmo antes de me conhecer, ter respondido meus e-mails, realizado sessões de Skype, ter disponibilidade em me encontrar em Liubliuana (Eslovênia) e ter respondido meus questionamentos quanto ao trabalho de organização do MulTeC enquanto o estava elaborando, sempre procurando esclarecer todas as minhas dúvidas.

Ao Prof. Dr. Rodrigo Esteves de Lima-Lopes cuja contribuição no debate SELIN de meu trabalho suscitou reflexões e contribuiu bastante para a organização do MulTeC.

Às Prof.^{as} Dr.^{as} Suzi Spatti Cavalari e Claudia Zavaglia que pacientemente apontaram as inúmeras falhas do texto apresentado na qualificação. Sua leitura cuidadosa e a discussão no dia da qualificação foram muito importantes para a continuidade do trabalho.

À Prof.^a Dr.^a Fernanda Galli pela orientação na qualificação especial. Suas palavras de encorajamento e sua competência conduziram-me a conclusão do trabalho quando imaginei que não conseguiria.

À Prof.^a Dr.^a Fabiana Komesu que, mesmo envolta em tantas atribuições, foi durante esses 4 anos uma presença amiga, um sorriso encorajador e uma capacidade de me convidar a realidade que foram especiais para mim.

À Ana Carolina Freschi, pela amizade, pelas revisões em minhas traduções Francês – Português e pela aula instrumental de Francês, além das conversas sempre tão sinérgicas e carinhosas.

À Ludmila Andreu-Funo pela leveza no compartilhar tantos anseios comigo, pelas leituras carinhosas e apontamentos questionadores ao meu trabalho.

À Deise Marinoto, pela tradução, feita tão prontamente, do artigo de Anne-Laure Foucher com profissionalismo e gentileza.

Aos meus filhos, Agatha Lima Lopes e Eduardo Lima Lopes, e esposo Francisco S. de Lima Silva, que estiveram apoiando minhas decisões, permanecendo carinhosamente ao meu lado, aceitando tantas mudanças, algumas restrições e o enfrentamento de algumas dificuldades em decorrência desse processo de doutoramento tão longe de nosso lar.

À minha mãe, Francisca S. Barbosa cuja paciência só aumenta e a capacidade de amar se expande com o passar dos dias. Agradeço por cuidar dos meus filhos, Agatha e Eduardo, por alguns meses para que pudesse organizar o retorno deles para nossa terra com mais tranquilidade.

Às minhas irmãs Clícia B. Lopes, Jardane B. Lopes e Betânia B. Lopes, pelo apoio incondicional sempre. Quanto amor se pode enviar virtualmente e fortalecer uma pessoa. Minha gratidão!

À minha querida amiga Lori Daiane Nunes Pereira, e seus dois filhos: João Asaph Fanini Costa Pereira e Samuel Fanini Costa Pereira, cuja luz, doçura e presenças cheias de amor foram fundamentais nos últimos dias de escrita da tese quando não pude estar mais perto dos meus filhos e esposo.

À minha querida amiga Laura Rampazzo que sempre esteve muito próxima a mim, intercedendo, enviando mensagens diárias de apoio quando estive nos piores dias da depressão, me ouvindo pacientemente sempre que precisei, lendo e relendo os meus textos, procurando me fazer acreditar em minha capacidade.

Às amigas do grupo de pesquisa teletandem de São José do Rio Preto, pelo compartilhamento de sorrisos, lágrimas e discussões sempre tão frutuosas.

Às professoras Ana Claudia Paleta e Daniela Águas que com suas aulas de natação e yoga, respectivamente, foram indispensáveis para o bem-estar necessário para concluir este trabalho.

Aos queridos psicólogos Lucas Cucato Bini e Daniella Fernandes Almeida que, com sua escuta qualificada e sempre tão respeitosa às minhas limitações e lágrimas, foram luz e reparação quando a depressão parecia que vencer minha capacidade de seguir.

Ao Prof. Dr. Júlio César Araújo cujo exemplo sempre me inspirou, seja pela escrita acadêmica, seja pelas palestras e cursos, a ser uma pessoa e uma pesquisadora melhor. Ter sua amizade desde 1997 é motivo de gratidão imensa, além de ser oportunidade de aprendizagem constante tanto no sentido acadêmico quanto espiritual e pessoal. Sei que ainda tenho muito a aprender.

À minha amiga Raquel Ishii que se fez presente durante esses quatro anos longe de meu lar, apesar de suas inúmeras atribuições. Cada e-mail afetuoso, cada presente recebido, cada gesto de amor inesperado nutriu meu coração e foi um refrigério para minha alma.

À família de Paulo e Michelle Arêas que acolheu a mim e a minha família com tanta generosidade e amor.

A tantos amigos feitos durante a pós-graduação. Vocês me constituem e me formam, porque, apesar de precisarmos nos posicionar e decidir como sujeitos, o fazemos a partir do que vivemos juntos como comunidade. Minha eterna gratidão!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) - código de financiamento 001, à qual agradeço.

RESUMO

O teletandem (TTD), contexto de aprendizagem de línguas telecolaborativo e autônomo (TELLES, 2006), desenvolvido na Unesp há mais de 12 anos, vem promovendo oportunidade de aprendizagem de línguas estrangeiras embasada na autonomia, separação de línguas e reciprocidade, princípios basilares da aprendizagem *in tandem*, como discutido por Brammerts (1996). Tal contexto de aprendizagem se tornou possível pela utilização das redes telemáticas que viabilizam a aprendizagem de línguas entre aprendizes que residem em países geograficamente distantes. No TTD, os aprendizes realizam sessões orais de teletandem (SOTs) e participam de sessões de mediação. Segundo Aranha e Leone (2016; 2017) essas são macrotarefas a partir das quais ocorrem as microtarefas: produção de diários de aprendizagem, troca de textos entre os participantes, respostas aos questionários inicial e final e uso da ferramenta chat durante a SOT. Esse contexto propicia a geração de uma quantidade significativa de dados multimodais (áudio, vídeo, texto escrito) que podem ser utilizados para pesquisas diversas. O intuito desta pesquisa é a construção de um *corpus* com os dados produzidos por 112 pares no período de 2012 a 2015. Para a composição do *corpus*, foram utilizados os dados coletados e armazenados conforme descrito por Aranha, Luvizari-Murad e Moreno (2015) e para a elaboração do *corpus* utilizei os seguintes procedimentos: a) levantamento das características de cada grupo de teletandem; b) criação de código para identificação dos parceiros; c) criação de códigos para a nomeação dos dados; d) padronização da anonimização dos dados; e) conversão dos dados para formatos compatíveis com diferentes sistemas operacionais; e f) renomeação dos dados a partir de Chanier e Wigham (2016). O MulTeC (Multimodal Teletandem *Corpus*) é composto por um total de 91 questionários iniciais, 41 questionários finais, 666 diários de aprendizagem, 355 chats, 921 textos produzidos na troca de texto por brasileiros e corrigidos por estadunidenses e 581 horas de gravações em vídeos, com transcrições em andamento. Após a finalização do *corpus*, percebeu-se a necessidade de ampliação de dados a serem coletados, incluindo a coleta da reunião tutorial e das sessões de mediação em mídia de vídeo ou áudio.

Palavras-chave: Teletandem. *Corpus*. Tarefas.

ABSTRACT

Teletandem (TTD), a telecollaborative and autonomous language learning context (TELLES, 2006), which has been carried out at Unesp for more than 12 years, has been promoting language learning based on autonomy, separation of languages and reciprocity, basic principles of tandem learning, as discussed by Brammerts (1996). Such learning context has become viable using telematic networks that enable language learning between learners residing in geographically distant countries. In teletandem, learners hold teletandem oral sessions (TOSs) and participate in mediation sessions. According to Aranha and Leone (2016, 2017) these are macrotasks from which microtasks occur: production of learning diaries, text exchange among TTD participants, responses to the initial and final questionnaires, and use of the chat tool during teletandem oral sessions. The production of these tasks generates a significant amount of data that can be used for various researches. The purpose of this research is constructing a multimodal *corpus* in teletandem with the data produced by the participants in the period from 2012 to 2015. The data which constitute the *corpus* were produced by 112 pairs, in a partnership with an American university in the period of 2012 to 2015, whose collection was described by Aranha, Luvizari-Murad and Moreno (2015). The elaboration of the *corpus* from the collection and storage described in Aranha, Luvizari-Murad and Moreno (2015) adopted the following procedures: a) survey of the characteristics of each teletandem group; b) creation of code to identify the partners; c) creation of codes for data naming; d) standardization of data anonymization; e) converting the data to compatible formats with different operating systems; and f) renaming the data according to what is proposed by Chanier and Wigham (2016). After the organization of the *corpus*, MulTeC (Multimodal Teletandem *Corpus*) has a total of 91 initial questionnaires, 41 final questionnaires, 666 learning diaries, 355 chats, 921 texts produced in the text exchange between the partners and 581 hours of TOSs recordings, with transcriptions ongoing. After the completion of the *corpus*, it was perceived the need to broaden the collection of the data, including tutorial meeting and mediation sessions in audio or video media collection.

Keywords: Teletandem. *Corpus*. Tasks.

LISTA DE ILUSTRAÇÕES

<u>Figura 1 – As modalidades de teletandem adaptado de Zakir (2015)</u>	23
<u>Figura 2 – Organização do cenário de aprendizagem no TTD por Aranha e Leone (2017)</u>	29
<u>Figura 3 – Versão dinâmica dos níveis de intervenção didática em Foucher (2010)</u>	33
<u>Figura 4 – Composição da macrotarefa por Grosbois (2009)</u>	35
<u>Figura 5 – Tarefas realizadas para mediação por objetivos</u>	36
<u>Figura 6 – Relação das tarefas na constituição do TTD</u>	38
<u>Figura 7 – Novo ciclo de vida dos dados</u>	47
<u>Figura 8 – Mapa dos dados</u>	49
<u>Figura 9 – Quadro comum europeu apresentado no questionário inicial do TTD</u>	70
<u>Figura 10 – Exemplo de anonimização nos textos revisados</u>	79
<u>Figura 11 – Cabeçalho do pareamento de turma TTD de 2013</u>	85
<u>Figura 12 – Cronologia no TTD</u>	89
<u>Figura 13 – Documento Anonimização dos Dados</u>	90
<u>Figura 14 – Documento ‘Código de Levantamento de Dados’</u>	91
<u>Figura 15 – Código Informações Interagentes</u>	92
<u>Figura 16 – Indicação de participante recorrente</u>	94
<u>Figura 17 – Aba ‘Dados produzidos por par’ – Planilha Levantamento de Dados</u>	96
<u>Figura 18 – Abertura de pasta de dados produzidos por par</u>	99
<u>Figura 19 – Aba ‘Contagem de palavras’ na planilha ‘Levantamento de Dados’</u>	101
<u>Figura 20 – Planilha ‘Levantamento de Dados Geral’</u>	105
<u>Figura 21 – Planilha ‘Levantamento de Dados Geral’</u>	107
<u>Figura 22 – Texto produzido por participante</u>	112
<u>Figura 23 – Representação da organização das pastas no MulTeC</u>	114
<u>Figura 24 – Cenário pedagógico – Aba 1</u>	117
<u>Figura 25 – Cenário pedagógico – Aba 3</u>	119
<u>Figura 26 – Cenário de aprendizagem – Aba 2</u>	121
<u>Figura 27 – Cenário de aprendizagem – Aba 3</u>	121
<u>Figura 28 – Cenário de aprendizagem – Aba 4</u>	122
<u>Figura 29 – Protocolo de coleta apresentado ao grupo TTD de São José do Rio Preto</u>	124
<u>Figura 30 –Relação entre os procedimentos de coleta, desafios e ciclo de dados</u>	127

LISTA DE QUADROS

<u>Quadro 1 – Relação entre as perguntas feitas na análise da organização do Banco TTDii e suas respectivas descobertas</u>	58
<u>Quadro 2 – Relação entre procedimentos e objetivos específicos</u>	63
<u>Quadro 3 - Síntese dos questionamentos, procedimentos e fontes</u>	71
<u>Quadro 4 – Protocolo de criação do IT.....</u>	76
<u>Quadro 5 – Padronização da nomenclatura dos arquivos</u>	83
<u>Quadro 6 – Exemplo de nomeação de dados do participante I8F7 (UGA3i – 2015).....</u>	84
<u>Quadro 7 – Sistema de Transcrição para o MulTeC.....</u>	102

LISTA DE ABREVIATURAS E SIGLAS

TTD – Teletandem

TTDii – Teletandem institucional integrado

TTDisi – Teletandem institucional semi-integrado

SOT – Sessão Oral de Teletandem

SOTi – Sessão Oral de Teletandem inicial

SOTin – Sessão Oral de Teletandem intermediária

SOTf – Sessão Oral de Teletandem final

DOTI – Databank of Oral Teletandem Interactions

IS – Interaction Space

LC – Linguística de *Corpus*

IoT – Internet of Things

TCLEs – Termos de Consentimento Livre e Esclarecido

TICEs – Tecnologias de Informação e Comunicação na Escola

CETIC – Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação

SUMÁRIO

INTRODUÇÃO	14
CAPÍTULO 1: EMBASAMENTO TEÓRICO	19
1.1 TELETANDEM: TELECOLABORAÇÃO PARA APRENDIZAGEM DE LÍNGUAS ESTRANGEIRAS.....	19
1.1.1 AS MODALIDADES DE (TELE)TANDEM	21
1.1.1.1 <i>Teletandem institucional: integrado, não-integrado e semi-integrado</i>	24
1.2 O TELETANDEM E A PRODUÇÃO DE DADOS: CONCEITOS IMPORTANTES.....	27
1.2.1 CENÁRIOS DE APRENDIZAGEM/PEDAGÓGICO E A RELAÇÃO ENTRE MACROTAREFAS, TAREFAS E MICROTAREFAS	32
1.2.1.1 RELAÇÃO ENTRE MACROTAREFA, TAREFAS E CENÁRIO DE APRENDIZAGEM NO TTD.....	36
1.3 A LINGÜÍSTICA DE <i>CORPUS</i>	39
1.3.1 COMPREENSÃO DA LINGÜÍSTICA DE <i>CORPUS</i> E TIPOLOGIA DE <i>CORPUS</i>	40
1.3.2 <i>CORPUS</i> MULTIMODAL: A NOVA GERAÇÃO DOS CORPORA	41
1.4 O GERENCIAMENTO DE DADOS PARA CONSTRUÇÃO DE <i>CORPUS</i> DE PESQUISA.....	46
1.4.1 O PLANO DE GERENCIAMENTO DOS DADOS	50
1.4.2 OS METADADOS NO DOTI (DATABANK OF ORAL TELETANDEM INTERACTIONS)	53
CAPÍTULO 2: PERCURSOS METODOLÓGICOS	55
2.1 ABORDAGEM DA PESQUISA.....	55
2.2 CONTEXTUALIZANDO A PESQUISA	57
2.2 AS 3 ETAPAS PARA A CONSTRUÇÃO DO <i>CORPUS</i>	61
2.2.1 UM OBJETIVO: VÁRIOS PROCESSOS	62
2.2.1.1 <i>O início do processo de organização: coleta de informações sociolinguísticas</i>	63
2.2.1.1.1 O TCLEs	66
2.2.1.1.2 Proficiência dos participantes	69
2.2.1.1.3 Documentos por turma TTD no Banco TTDii	72
2.2.1.2 <i>Procedimentos para o tratamento dos dados: anonimização, conversão e nomeação</i>	75
2.2.1.2.1 Definição do código de identificação do participante	76
2.2.1.2.2 Anonimização dos dados	77
2.2.1.2.3 Conversão dos dados	81
2.2.1.2.4 A padronização da nomeação dos arquivos	82
2.2.1.3 <i>Organização dos arquivos no MulTeC: etapa 3 da organização do MulTeC</i>	85
2.2.1.3.2 Documentos com códigos utilizados na constituição do <i>corpus</i>	86
2.2.2 <i>Documentos contextuais</i>	88
2.2.2.1 Contexto TTD	88
2.2.2.2 Documentos com Códigos utilizados na nomeação e anonimização	89

2.2.2.3 Códigos das planilhas	90
2.2.3 PLANILHAS.....	92
2.2.3.1 Planilha ‘Informações participantes’	93
2.2.3.2 Planilha ‘Levantamento de Dados’	95
2.2.3.2.1 Levantamento de dados – aba 1.....	95
2.2.3.2.2 Contagem das palavras – aba 2.....	100
2.2.3.3 Planilha ‘Levantamento Geral’	104
2.2.4 O CICLO DOS DADOS NO MULTEC	106
2.2.5 CABEÇALHOS.....	108
2.2.6 CONVERSÃO DOS DADOS	110
2.3 TIPOLOGIA DO MULTEC.....	112
CAPÍTULO 3: RESULTADOS E DISCUSSÃO.....	113
3.1 RESULTADOS: O MULTEC E A PROPOSTA DE COLETA PARA O TTD.....	113
3.1.1 A ORGANIZAÇÃO DAS PASTAS	113
3.5 DOCUMENTOS PROPOSTOS: CENÁRIOS	115
3.6 A PROPOSTA DOS CENÁRIOS	116
3.6.1 CENÁRIO PEDAGÓGICO	116
3.6.2 CENÁRIO DE APRENDIZAGEM.....	120
3.7 A PROPOSTA DE COLETA AO TTD	123
CONSIDERAÇÕES FINAIS.....	129
OBJETIVOS E RESULTADOS ALCANÇADOS	129
LIMITAÇÕES DA PESQUISA.....	130
ENCAMINHAMENTOS FUTUROS	130
REFERÊNCIAS	133
Anexo 1 – Lista de presença	137
Anexo 3 – Tutorial.....	140
Anexo 4 - Tabela de Cursos Unesp.....	144
Anexo 5 - Contexto TTD	145
Anexo 6 – Nomeação dos dados.....	155
Anexo 7 – Plano de gerenciamento de dados - MulTeC	159
Anexo 8 – Organização do MulTeC.....	161

INTRODUÇÃO

“O trabalho colaborativo, a socialização em comunidades de pessoas que se reúnem por afinidade e por grupos de interesses comuns, são a tônica dos ambientes digitais” (COSCARELLI, 2016, p. 13)

Em tempos nos quais se discute a chegada da web 3.0, ou web semântica, que sinaliza para uma navegação muito mais significativa e produtiva, bem como a internet das coisas (IoT – *Internet of Things*), a utilização da internet parece cada dia mais fazer parte da vida da maioria dos seres humanos, como presença quase inevitável em seu cotidiano. No Brasil, segundo pesquisa realizada pelo Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação (Cetic.br¹), em 2016, 93% dos domicílios brasileiros tinham celulares e, em 79% deles, havia possibilidade de acesso à internet por quaisquer de seus moradores a qualquer momento. Tal cenário possibilita que pessoas geograficamente distanciadas possam identificar aqueles que têm os mesmos propósitos e, então criar grupos que nascem do compartilhamento desses objetivos, realizar trabalhos juntos, partilhando e gerando conhecimento colaborativamente via ambientes digitais. Esses ambientes, aos quais se refere Coscarelli (2016) na epígrafe, além de facilitarem formação de comunidades por compartilhamento de propósitos, ainda oportunizaram a emergência de um movimento na Ciência conhecido como *Open Science* (Ciência Aberta, em tradução livre).

Segundo Albagli (2015) *Open Science* se “[...] torna um termo abrangente que vai além do livre acesso a publicações científicas e inclui outros tópicos como dados científicos abertos, ferramentas científicas abertas [...]”² (ALBAGLI, 2015, p. 14-15). No mesmo sentido, Delfanti e Pitrelli (2015, p. 59) afirmam que o termo “é um conceito amplo que engloba diferentes práticas e ferramentas relacionadas ao uso colaborativo de tecnologias digitais e ferramentas de propriedade intelectual alternativas”³. Desse modo, ciência aberta se faz em todas as fases da pesquisa científica incluindo a produção e organização de dados para pesquisa de modo que sejam compartilháveis, assim como a publicação de resultados em mídias de livre acesso. *Open*

¹ <http://data.cetic.br/cetic/>

² No original: “Open Science becomes an umbrella term that goes beyond free access to scientific publications and includes other topics such as open scientific data, open scientific tools” (Tradução feita pela autora desta tese)

³ No original: “is a very broad concept that encompasses several different practices and tools linked to the use of collaborative digital technologies and alternative intellectual property tools”. (Tradução minha)

Science é, portanto, um paradigma de ciência no qual os pesquisadores realizam seus trabalhos pensando no compartilhamento não apenas do produto final, por meio das publicações de seus resultados, mas viabilizando também o compartilhamento dos dados oriundos do processo da pesquisa, das anotações feitas e ferramentas utilizadas.

A rede mundial de computadores ainda tem oportunizado contexto de intercâmbio virtual⁴ para aprendizagem de línguas a partir do uso das ferramentas de comunicação síncrona disponíveis na rede mundial de computadores, como por exemplo a aprendizagem telecolaborativa via teletandem.

Tanto *Open Science*, como a criação de grupos de interesse de qualquer natureza em ambientes digitais e o intercâmbio virtual, apesar de serem de níveis diferentes, podem ser tomados como exemplos de compartilhamento à distância que foram tornados mais céleres e amplos a partir da *web 2.0*. Segundo Primo (2007, p. 2), a internet dessa geração “caracteriza-se por potencializar as formas de publicação, compartilhamento e organização de informações, além de ampliar os espaços para a interação entre os participantes do processo”.

A formação de comunidades geograficamente distantes, agregadas virtual e sincronicamente só se tornou possível por meio da rede mundial de computadores, pois a comunicação por telegramas é uma possibilidade bem anterior a ela, mas seu caráter assíncrono a diferencia do uso da web. Essas comunidades são movidas pelo compartilhamento de propósitos com trocas mais fluídas e em menos tempo que as tecnologias anteriores como o telefone e o telegrama. Segundo Silva (2012, p. 85), o TTD é uma comunidade que “se estabelece a partir de afinidades de interesses, objetivos em comuns, partilha de um mesmo conhecimento e de uma prática”. Essa comunidade existe pelo desejo que pares interagentes têm de que ocorra aprendizagem de uma língua estrangeira. Assim, esse propósito une seus participantes em um contexto digital que desconhece fronteiras geográficas e torna factível o intercâmbio cultural tão amplo e facilitado como inimaginável até algumas décadas atrás.

No contexto TTD, participantes proficientes em línguas diferentes se dispõem a serem tutores um do outro nas respectivas línguas-alvo, de modo que, por exemplo, o participante proficiente em português auxilia a aprendizagem do parceiro, proficiente em inglês – língua-alvo do primeiro. Os participantes devem respeitar os princípios basilares da aprendizagem *in*

⁴ Gostaria de destacar que essa terminologia – intercâmbio virtual – é utilizada por O’Dowd (2018) para nomear o “engagement of groups of learners in *on-line* intercultural interactions and collaboration projects with partners from other cultural contexts or geographical locations as an integrated part of their educational programmes (p. 1)”

tandem inicialmente propostos por Brammerts (1996), de autonomia e reciprocidade, mantidos por Telles (2006), como é possível perceber no excerto em que afirma que a decisão sobre “o como” e “o que” irão aprender deve ser feita “em comum acordo por ambos os participantes, em consonância com um dos dois princípios da noção de tandem – a “autonomia dos aprendizes” (o outro princípio é a “reciprocidade”)” (TELLES, 2006, p. 10). No entanto, há ainda um terceiro princípio, a separação de línguas. Desse modo, no TTD, os princípios são três, a saber: autonomia, reciprocidade e separação de línguas. Seguindo esses princípios, os participantes devem: i) sob o princípio da autonomia, estabelecer seus objetivos de aprendizagem e o modo como desejam aprender; assim como ii) respeitar as necessidades de aprendizagem do parceiro, sendo-lhe recíproco (reciprocidade) e iii) dividir igualmente o tempo dedicado às línguas que objetivam aprender durante a parceria (separação de línguas).

Além de ser projeto de aprendizagem de línguas, o TTD também se caracteriza como projeto de pesquisa, envolvendo desde 2006, ano de início de suas atividades, até o presente momento, diversos pesquisadores que utilizaram (e ainda utilizam) os dados produzidos pelos participantes para investigações científicas. Cada pesquisador interessado em utilizar os dados produzidos no TTD fazia sua coleta a partir de suas perguntas de pesquisa, organizando os dados de seu estudo conforme protocolos estabelecidos individualmente, o que implicava em uma multiplicidade de formas de sistematização de material documentário e uma consequente impossibilidade de sistematização de um banco de dados mais abrangente, capaz de favorecer futuros estudos⁵.

A partir de 2011, conforme Aranha, Luvizari-Murad e Moreno (2015) o grupo de pesquisadores em TTD da Unesp de São José do Rio Preto começou a realizar coleta dos dados produzidos pelos participantes para futuras pesquisas, a serem utilizados por pesquisadores do TTD. Assim, as atividades realizadas entre os participantes, a saber: as interações via vídeo, os chats, os textos produzidos para revisão pelo parceiro, os diários de aprendizagem (produzidos para ajudar na reflexão sobre a aprendizagem via TTD) e os questionários iniciais e finais, foram sendo coletados e organizados conforme protocolo estabelecido pelo grupo de pesquisa e detalhado por Aranha, Luvizari-Murad e Moreno (2015), criando o Banco TTDii - Banco de Dados para pesquisas sobre aprendizagem via Teletandem institucional integrado⁶ (TTDii).

⁵ Informações mais detalhadas sobre pesquisas realizadas no TTD podem ser obtidas <https://teletandemriopreto.wixsite.com/ibilce/pesquisas>

⁶ O teletandem institucional integrado é uma modalidade possível no TTD. Discorro sobre este trabalho mais detalhadamente sobre as modalidades de TTD no capítulo de fundamentação teórica.

Em síntese, uma quantidade significativa de dados foi salva e armazenada, contudo, os dados não estavam de acordo com o previsto para que pudessem ser compartilháveis, ou seja, havia necessidade de anonimização, padronização da nomenclatura e elaboração de documentos contextuais. A dimensão do problema engendrado pela ausência de procedimentos relevantes para tornar os dados compartilháveis, evidencia-se especialmente, quando consideramos que o compartilhamento possibilitaria que pesquisadores das mais diversas áreas pudessem acessá-los, contribuindo ainda mais para o desenvolvimento da ciência.

Desse modo, a presente pesquisa emerge da necessidade de potencializar o alcance do relevante trabalho realizado pelo grupo, tornando o Banco TTDii em um *corpus*. Assumindo essa missão e partindo da hipótese de que seria possível construir um *corpus* a partir dos dados coletados, estabeleci as seguintes perguntas de pesquisa:

- Quais procedimentos precisam ser adotados para que o banco de dados TTDii possa se constituir um *corpus* cujos dados possam ser compartilháveis com pesquisadores que não participaram da coleta?
- Quais informações precisam ser acrescentadas?
- Quais documentos precisam ser criados para auxiliar o uso dos dados em pesquisas?

Ao responder essas perguntas alcanço o objetivo geral desta pesquisa, a saber:

- Construir um corpus multimodal em teletandem com as produções coletadas no contexto TTD da Unesp de São José do Rio Preto, a partir do levantamento, anonimização e organização dos dados produzidos no período de 2012 a 2015.

Para alcançar esse objetivo, defini como objetivos específicos os seguintes:

- Analisar a organização dos dados produzidos e coletados no período de 2012 a 2015, conforme descrito em Aranha, Luvizari-Murad e Moreno (2015);
- Propor padrão de anonimização e nomenclatura dos arquivos (BRINEY, 2015; CHANIER; WIGHAM, 2016; EYNDEN et al., 2011);
- Organizar e analisar os dados a partir dos pressupostos teóricos de organização de dados de pesquisa (BRINEY, 2015; CHANIER; WIGHAM, 2016; EYNDEN, 2013; EYNDEN et al., 2011), elaborando e acrescentando documentos para a construção de um corpus de pesquisa.

Esta pesquisa então apresenta, portanto, um *corpus* a ser compartilhado. Esta tese, que discute a construção desse *corpus*, tem as informações distribuídas em três capítulos. No primeiro capítulo, apresento o embasamento teórico utilizado para a realização do trabalho,

mais especificamente, partindo do conceito de telecolaboração e do contexto TTD de aprendizagem de línguas, discutindo, em seguida, o conceito de tarefas assim como a constituição de *corpus* a partir dos pressupostos da Linguística de *Corpus* (LC) e dos protocolos internacionais para a construção de *corpus* pesquisa.

No segundo capítulo, discorro sobre a metodologia utilizada na organização do *Multimodal Teletandem Corpus* (MulTeC), explicitando os procedimentos e padrões adotados a partir das leituras de trabalhos que organização de *corpus* como os de Chanier e Wigham (2016), Briney (2015) e Eynden *et al.* (2011), bem como os resultados obtidos a partir da organização dos dados.

No capítulo três, apresento os resultados e uma discussão a partir do que experienciei durante a organização do MulTeC, no qual apresento ainda a proposta para coleta e organização dos dados produzidos e coletados no contexto de aprendizagem de línguas TTD. Encerro esta tese com as considerações finais com a retomada dos objetivos de pesquisa, discussão sobre as limitações da pesquisa e encaminhamentos futuros.

O percurso discursivo se inicia com os embasamentos teóricos que viabilizaram um olhar sobre os dados produzidos, seu contexto de produção e estratégias de gerenciamento de dados que otimizem o tempo dedicado a coleta e organização.

CONSIDERAÇÕES FINAIS

Nesta parte do trabalho, apresento uma breve retomada dos objetivos que conduziram a realização desta pesquisa, passando então a discutir as limitações do trabalho e encerrando com os encaminhamentos futuros a partir do que foi detectado durante o processo de realização da pesquisa e escrita deste texto.

OBJETIVOS E RESULTADOS ALCANÇADOS

Iniciei este trabalho com a hipótese de que seria possível construir um *corpus* multimodal em teletandem com as produções coletadas no TTD da Unesp de São José do Rio Preto, a partir do levantamento, anonimização e organização dos dados produzidos no período de 2012 a 2015. Para alcançar a realização desse objetivo, precisei dos seguintes objetivos específicos: a) investigar a organização dos dados; b) padronizar a nomeação e anonimização e c) organizar e analisar os dados. No percurso houve necessidade de seguir os padrões que estabeleci quando da elaboração do plano de gerenciamento dos dados. No plano de gerenciamento de dados (Anexo 7), defini como os dados deveriam ser organizados e anonimizados, estabelecendo os padrões de nomeação, inclusive.

Cito como resultados, além do próprio MulTeC, sua proposta de organização de dados para o TTD, a proposta de uso de documentos que representem o cenário pedagógico e o cenário de aprendizagem. Esses documentos se configuram como relevantes para a organização de *corpus* de coleta no TTD da Unesp de São José do Rio Preto, pois documentos fornecerão informações indispensáveis para a compreensão do contexto de produção dos dados. Como informações necessárias para a construção do contexto no *corpus* tais documentos devem apresentar: i) a modalidade, ii) a turma, iii) os professores e os mediadores, iv) registro das ocorrências em cada SOT; v) registro sobre fatos ocorridos durante a coleta que possam influenciar na apresentação dos dados no *corpus*; e vi) nome da turma, que seguindo os padrões estabelecidos para o MulTeC, vai informar a modalidade de TTD na qual a turma está ocorrendo.

Os dados indicaram que os tipos de dados produzidos no teletandem dependem mais das decisões tomadas durante o planejamento da turma (cenário pedagógico) do que da modalidade na qual o mesmo ocorre.

LIMITAÇÕES DA PESQUISA

Encontrei entraves que impediram a conclusão da construção do MulTeC, conforme as intenções expressas em meus projetos iniciais. Elenco como percalços: i) o tempo, considerando que transcrever mais que 500 horas de vídeo não pode ser realizado no âmbito temporal de um doutorado e ii) a inexistência de algumas informações, sobretudo no tocante a ocorrências durante as SOTs, haja vista que não houve registro na maioria das turmas, o que levou, por exemplo a exclusão de um ano de coleta de dados realizada em 2011.

A ausência de documentos com registros sobre o contexto de produção mais detalhado tornou-se um obstáculo para o preenchimento de todas as informações das instituições constituintes da parceria em algumas planilhas, como no caso da planilha que apresenta as informações sobre os participantes (idade, gênero, curso), pois, no que se refere aos participantes da universidade estrangeira que fez parceria com a Unesp, cujos dados foram utilizados na construção do banco, não foi possível coletar informações sobre o curso e a idade, pois não havia documento com tal registro.

ENCAMINHAMENTOS FUTUROS

Os dados sugerem a necessidade de ampliação dos dados incluindo mais tipos de dados, como a reunião tutorial, que pode ser gravada em áudio ou vídeo, bem como as sessões mediação. Essa necessidade advém da compreensão de que ambos constituem tanto o cenário pedagógico quanto o de aprendizagem, podendo influir nas produções dos participantes no TTD. Ainda apresento como sugestão, a partir deste trabalho, uma ampliação do questionário inicial. Entendo que esse documento pode servir para coletar informações sociolinguísticas que contribuirão na construção do perfil sociolinguístico do participante sem descuidar da preservação de sua identidade, podendo haver a inserção da informação de tempo de estudo da língua alvo, por exemplo, bem como quantas línguas fala/estuda. No contexto do TTD questionar apenas sobre L1 e L2 me parece um tanto quanto reduzido considerando que, nos dados utilizados para este trabalho, mais especificamente nas SOTs que precisei realizar oitiva, identifiquei falantes de língua inglesa que informavam aos parceiros que sabiam falar espanhol além do português, língua-alvo para eles na parceria em questão. Outro levantamento relevante é a idade do participante, além da data de nascimento, para que aqueles que estejam responsáveis pela coleta e organização dos dados não precisem realizar o cálculo, como precisamos fazer para este trabalho, considerando que esse é um fator relevante na análise dos

dados pelos pesquisadores que podem, inclusive, optar por selecionar dados produzidos por um faixa etária definida.

Considerando que a participação dos aprendizes pode ocorrer em mais de um semestre, como detectei nos dados, o ideal na anonimização seria a utilização de um IT único por participante. Ao participar pela primeira vez do contexto TTD seria criado o IT seria utilizado em todos os dados produzidos por ele quando participasse do TTD. A pergunta no questionário inicial sobre a participação em oportunidade anterior no teletandem seria o ponto de partida, para a equipe responsável pela coleta, procurar o IT criado para aquele participante e então nomear os dados produzidos por ele com o mesmo IT. Uma proposta de IT único poderia ser, por exemplo, constituída pelo ano que participou pela primeira vez, instituição, semestre, curso, usuário Skype e gênero declarado: 2018I127S13F – participante declarada feminina, cursando o doutorado em estudos linguísticos, participou pela primeira vez em 2018 como aluna do Ibilce, no primeiro semestre, utilizando o usuário Skype 13. Para registro e para que sequenciação dos ITs criados tenha consistência, o ideal seria a elaboração e manutenção de uma planilha com os ITs criados com: nome, curso, data de nascimento e e-mail do participante de modo que possa ser facilmente localizado quando for realizar TTD novamente.

A análise dos dados ainda aponta para a necessidade de uma equipe responsável pela coleta, havendo pelo menos alguém da área de TI e uma coordenadora do processo de coleta e organização dos dados no intuito de manter a consistência do padrão de anonimização e nomeação dos dados. O papel dessa pessoa, que estará responsabilizada pela compreensão do plano de gerenciamento de dados e por quem estiver responsável pela coleta e organização, deveria consistir na orientação e revisão de cada dado coletado, observando se está se seguindo o padrão convencional, bem como procurar dirimir dúvidas durante o processo de gerenciamento de dados por aqueles que estão responsáveis pela coleta.

Outra questão relevante a ser considerada é que os documentos produzidos para esta pesquisa, como as planilhas, cenário pedagógico e de aprendizagem¹⁰², não terão qualquer utilidade se não forem criteriosamente preenchidos pelas mediadoras/coordenadoras do TTD que integram a equipe de modo que se consiga levantar os dados contextuais indispensáveis para a construção de *corpus* de pesquisa que atenda aos pesquisadores interessados em investigar o TTD. Gostaria de destacar que os dados que constituem o MulTeC podem ser de

¹⁰² Foi disponibilizado para o grupo de pesquisa em TTD da Unesp de Rio Preto um modelo de cada um desses documentos. No caso das planilhas, já com as fórmulas, de maneira que quem for utilizá-las não precise dedicar tempo a esse trabalho, apenas precise inserir as informações coletadas.

interesse não apenas para cientistas que pesquisam a aprendizagem de línguas em projeto telecolaborativo, mas investigações que discutam:

- i) Interculturalidade
- ii) Identidade
- iii) Escrita de textos em língua estrangeira
- iv) *Feedback*
- v) Processo autorreflexivo da aprendizagem (a partir dos registros nos diários de aprendizagem)
- vi) Uso de ferramentas digitais para a intercompreensão;
- vii) Construção da aprendizagem autônoma.

Desse modo, acredito que envidar tais esforços deva contribuir para o fortalecimento das pesquisas em TTD em âmbito internacional, isso considerando o direcionamento que a ciência no mundo tem dado para o compartilhamento de dados de pesquisa objetivando tornar cada vez mais comum a *Open Science*.

REFERÊNCIAS

- ANDREU-FUNO, L. *Teletandem: um estudo sobre identidades culturais e sessões de mediação da aprendizagem*. Tese de Doutorado. Programa de Pós-Graduação em Estudos Linguísticos da Universidade Estadual Paulista “Júlio de Mesquita Filho” – São José do Rio Preto, 2015.
- ASSANTE, M.; CANDELA, L.; CASTELLI, D.; TANI, A.; *Are Scientific Data Repositories Coping with Research Data Publishing?* *Data Science Journal* 15, no. 6, p. 1-24., 2016. Disponível em: <http://doi.org/10.5334/dsj-2016-006> Acesso em: 12 de jul, 2017.
- ALBAGLI, S. Open science in question. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.) *Open Science, open issues*. Rio de Janeiro: Unirio, 2015. p. 9–22.
- ARANHA, S.; CAVALARI, S. M. S. A trajetória do Projeto Teletandem Brasil: Da modalidade institucional não-integrada à institucional integrada. *The ESpecialist*, v. 35, p. 183–201, 2014.
- ARANHA, S. Os gêneros na modalidade de teletandem institucional integrado: a primeira sessão de interação. *Gêneros de texto/discurso e os desafios da contemporaneidade*. [S.l.: s.n.], 2014. .
- ARANHA, S.; LEONE, P. DOTI: Databank of Oral Teletandem Interactions. *New directions in telecollaborative research and practice: selected papers from the second conference on telecollaboration in higher education*. [S.l.]: Research-publishing.net, 2016a. p. 327–332.
- ARANHA, S.; LEONE, P. DOTI: Databank of Oral Teletandem Interactions. *New directions in telecollaborative research and practice: selected papers from the second conference on telecollaboration in higher education*. [S.l.]: Research-publishing.net, 2016b. p. 327–332.
- ARANHA, S.; LEONE, P. The development of DOTI (Data of oral teletandem interaction). *Investigating Computer-Mediated Communication: Corpus-Based Approaches To Language in the Digital World*. Ljubljana: University of Ljubljana Press, Faculty of Arts, 2017. p. 172–190. Disponível em: <<https://e-knjige.ff.uni-lj.si>>.
- ARANHA, S.; LUVIZARI-MURAD, L. H.; MORENO, A. C. a Criação De Um Banco De Dados Para Pesquisas Sobre Aprendizagem Via Teletandem Institucional Integrado (Ttdii). *Revista (Con)textos Linguísticos*, v. 9, n. 12, p. 274–293, 2015.
- ARAUJO, N. R. P. *A formação de parcerias de teletandem: da organização ao sistema de atividade*. 2012. Universidade Estadual Paulista - UNESP/ Ibilce, 2012.
- BERBER SARDINHA, T. Pesquisa em linguística de *Corpus* com WordSmith Tools. *Mercado de Letras*, p. 299, 2009. Disponível em: <http://sis.posugf.com.br/AreaProfessor/Materiais/Arquivos_1/13879.pdf>.
- BERBER SARDINHA, T. Tamanho de *Corpus*. *the ESpecialist*, v. 23, n. 2, p. 103–122, 2002.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri: Editora Manole, 2004.
- BOGDAN, R. C.; BIKLEN, S. K. *Investigação qualitativa em educação - Uma introdução à teoria e aos métodos*. [S.l.: s.n.], 1994.

BRAMMERTS, H. Tandem language learning via internet and international e-mail tandem network. *A guide to language learning in tandem via internet*. [S.l.]: CLCS Occasional Paper, 1996. p. 9–22.

BRINEY, K. *Data management for researchers: Organize, maintain and share your data for research success*. Exeter: Pelagic Publishing, 2015.

CAVALARI, S.; ARANHA, S. Teletandem : integrating e-learning into the foreign language classroom. *Acta Scientiarum Language and Culture*, v. 38, n. November, p. 327–336, 2016.

CHANIER, T. *et al.* The CoMeRe corpus for French : structuring and annotating heterogeneous CMC genres. *Journal for Language Technology and Computational Linguistics*, v. 29, n. 2, p. 1–30, 2014.

CHANIER, T.; WIGHAM, C. R. A scientific methodology for researching CALL interaction data. *Language-learner computer interactions: Theory, methodology and CALL applications*, p. 215–240, 2016. Disponível em: <<https://benjamins.com/catalog/lse.2.10cha>>.

COSCARELLI, C. V. *Tecnologias para aprender*. 1 ed ed. São Paulo: Parábola Editorial, 2016.

DELFANTI, A.; PITRELLI, N. Open science: revolution or continuity? In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). *Open Science, open issues*. Rio de Janeiro: IBICT, 2015. p. 59–68.

DETOMINI, A. C. S. M. *Dúvidas Linguísticas de Aprendizizes de PLE: avaliação de proficiência oral de professores no EPPLE-PLE*. Dissertação de Mestrado. Programa de Pós-Graduação em Estudos Linguísticos. Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP, 2018.

DOOLY, M.; O’DOWD, R. Telecollaboration in the foreign language classroom : A review of its origins and its application to language teaching practice. In *This Together: Teachers’ experiences with transnational, telecollaborative language learning projects*. [S.l.: s.n.], 2018. p. 11–34. Disponível em: <<https://www.peterlang.com/view/9783034335348/chapter01.xhtml>>.

DÖRNEY, Z. *Research Methods in Applied Linguistics*. Oxford: [s.n.], 2009. v. 19. Disponível em: <<http://doi.wiley.com/10.1111/j.1473-4192.2009.00223.x>>.

ELLIS, R. *Task-based language learning and teaching*. New York: Oxford University Press, 2003.

EYNDEN, V. V. D. *Data Life Cycle & Data Management Planning*. UK Data Archive University of Essex, n. April, p. 24–25, 2013. Disponível em: <<https://www.ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf>>.

EYNDEN, V. V. D. *et al.* *Managing and Sharing Data - Best Practice For Researchers*. [S.l.: s.n.], 2011. Disponível em: <<http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>>.

FINATTO, M. J. *et al.* *Linguística de Corpus: perspectivas*. Porto Alegre: Instituto de Letras - UFRGS, 2018.

FOUCHER, A. *Didactique des Langues-Cultures et Tice : scénarios , tâches , interactions*. . [S.l: s.n.], 2010.

FRESCHI, A. C. *A avaliação por pares no teletandem institucional integrado: um estudo de caso sobre o feedback linguístico nas sessões orais em português*. Dissertação de Mestrado. Programa de Pós-Graduação em Estudos Linguísticos. Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP, 2017.

FRESCHI, A. C.; LOPES, Q. B. Potenciais Sequências De Aprendizagem Intercultural No Teletandem: a Importância Da Mediação. *Revista do GEL*, v. 13, n. 3, p. 49, 2016. Disponível em: <<https://revistadogel.gel.org.br/rg/article/view/1470>>.

GUICHON, N. *Langues et TICE: méthodologie de conception multimédia*. [S.l: s.n.], 2006.

GUICHON, N. Sharing a Multimodal *Corpus* To Study Webcam-Mediated Language Teaching. v. 21, n. 1, p. 56–75, 2017.

GÜNTHER, H. Pesquisa qualitativa versus pesquisa quantitativa: esta é a questão? *Psicologia: Teoria e Pesquisa*, v. 22, n. 2, p. 201–209, 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-37722006000200010&lng=pt&tlng=pt>.

KNIGHT, D. The future of multimodal corpora. *RBLA, Belo Horizonte*, v. 11(2), p. 391–415, 2011. Disponível em: <http://www.google.com/url?sa=t&rct=j&q=the future of multimodal corpora+pdf&source=web&cd=1&ved=0CFAQFjAA&url=http://www.periodicos.letras.ufmg.br/rbla/arquivos/332.pdf&ei=VRMIUJiJE4yU0QXLrqnzBA&usq=AFQjCNHeGzq-mj102Z2f9p_3yqEJDpaqNA&cad=rja>.

LUVIZARI-MURAD, L. H. *Aprendizagem de alemão e português via teletandem: um estudo com base na Teoria da Atividade*. 2011. Universidade Estadual Paulista (Unesp), Instituto de Biociências, Letras e Ciências Exatas (Ibilce), Câmpus de São José do Rio Preto, 2011.

MANGENOT, F. *La question du scénario de communication dans les interactions pédagogiques en ligne*. JOCAIR 2008 (Journées Communication et Apprentissage Instrumentés en Réseau), v. 2008, n. Ea 609, p. 13–26, 2008.

O'DOWD, R. *Emerging Trends and New Directions in Telecollaborative Learning*. CALICO Journal, v. 33, n. 3, p. 291–310, 2016.

PRIMO, A. *O aspecto relacional das interações na Web 2.0*. E-Compós: Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação, v. 9, p. 1–21, 2007. Disponível em: <<http://www.compos.org.br/seer/index.php/e-compos/article/view/153>>.

RAMPAZZO, L. *Gêneros textuais e telecolaboração : uma investigação da sessão oral de teletandem inicial*. Dissertação de Mestrado. Programa de Pós-Graduação em Estudos Linguísticos. Universidade Estadual Paulista “Júlio de Mesquita Filho” - UNESP, 2017.

REFFAY, C.; BETBEDER, M.; CHANIER, T. *Multimodal learning and teaching corpora exchange : lessons learned in five years by the Mulce project* Marie-Laure Betbeder Thierry Chanier. 2013.

SALOMÃO, A. C. B. *Gerenciamento e estratégias pedagógicas na mediação dos pares no*

teletandem e seus reflexos para as práticas pedagógicas dos interagentes. 2008. Universidade Estadual Paulista “Júlio de Mesquita Filho” - São José do Rio Preto, 2008.

SHEPHERD, T. M. G. *O Estatuto Da Linguística De Corpus: Metodologia Ou Área Da Linguística ? Matraca*, v. 16, n. 24, p. 150–172, 2009.

SINCLAIR, J. *Corpus and Text - Basic principles*. Disponível em: <<https://ota.ox.ac.uk/documents/creating/dlc/chapter1.htm>>. Acesso em: 20 out. 2017.

SOUZA, M.G. *Os primeiros contatos de professores de línguas estrangeiras com a prática de teletandem*. 2012. Universidade Estadual Paulista “Júlio de Mesquita Filho” - São José do Rio Preto, 2012. Disponível em: <<https://repositorio.unesp.br/handle/11449/93887>>.

TOLEDO, L. D. L. *Um estudo sobre o uso de vocabulário rico por aprendizes de inglês na sessão oral do teletandem institucional integrado*. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Estadual Paulista Júlio de Mesquita Filho, 2017.

WIGHAM, C. R.; CHANIER, T. *LEarning and TEaching Corpora (LETEC): data-sharing and repository for research on multimodal interactions*. Global perspectives on Computer-Assisted Language Learning, p. 10–13, 2013. Disponível em: <<https://hal.archives-ouvertes.fr/edutice-00778274/document>>.

ZAKIR, M. de A. *Cultura e(m) telecolaboração: uma análise de parcerias de teletandem institucional*. 2015. Universidade Estadual Paulista “Júlio de Mesquita Filho” - Campus de São José do Rio Preto, 2015.