

Júlia Vidigal Zara

To give you a book or to give a book to you: um estudo
sobre a variação sintática na expressão linguística de
eventos de transferência de posse na interlíngua
português/inglês

Belo Horizonte, Brasil

Faculdade de Letras, UFMG

2014

Júlia Vidigal Zara

To give you a book or to give a book to you: um estudo
sobre a variação sintática na expressão linguística de
eventos de transferência de posse na interlíngua
português/inglês

Texto apresentado ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Doutora em Linguística Teórica e Descritiva.

Área de Concentração: Linguística Teórica e Descritiva

Linha de Pesquisa: Estudos Linguísticos Baseados em Corpora

Orientadora: Profa. Dra. Heliana Ribeiro de Mello (UFMG)

Co-orientador: Prof. Dr. Crysttian Arantes Paixão (FGV)

Belo Horizonte, Brasil

Faculdade de Letras, UFMG

2014

Z36t Zara, Júlia Vidigal.
To give you a book or to give a book to you
[manuscrito] : um estudo sobre a variação sintática na
expressão linguística de eventos de transferência de
posse na interlíngua português / inglês / Júlia Vidigal
Zara. – 2014.
180 f., enc. : il., grafs, tabs, (color), (p&b)
Orientadora: Heliana Ribeiro de Mello.
Coorientador: Crysttian Arantes Paixão.
Área de concentração: Linguística Teórica e Descritiva.
Linha de Pesquisa: Estudos Lingüísticos Baseados em Corpora.
Tese (doutorado) – Universidade Federal de Minas
Gerais, Faculdade de Letras.
Bibliografia: f. 165-180.

1. Aquisição da segunda linguagem – Teses. 2.
Interlíngua (Aprendizagem de línguas) – Teses. 3.
Língua portuguesa – Estudo e ensino – Falantes
estrangeiros – Teses. 4. Linguística de *corpus* – Teses.
5. Língua inglesa – Variação – Teses. 6. Língua
inglesa – Sintaxe – Teses. I. Mello, Heliana Ribeiro de
. II. Paixão, Crysttian Arantes. III. Universidade
Federal de Minas Gerais. Faculdade de Letras. IV.
Título.

CDD: 420.7



FOLHA DE APROVAÇÃO

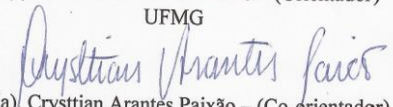
To give you a book or to give a book to you: um estudo sobre a variação sintática na expressão linguística de eventos de transferência de posse na interlíngua português/inglês

JÚLIA VIDIGAL ZARA

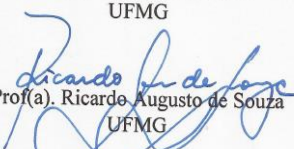
Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Linha G-Estudos Inter-Relação Linguagem, Cognição e

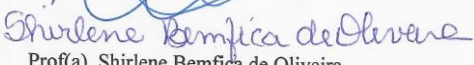
Aprovada em 17 de março de 2014, pela banca constituída pelos membros:


Prof(a). Heliana Ribeiro de Mello – (Orientador)
UFMG


Prof(a). Crysttian Arantes Paixão – (Co-orientador)
Fundação Getúlio Vargas


Prof(a). Deise Prina Dutra
UFMG


Prof(a). Ricardo Augusto de Souza
UFMG


Prof(a). Shirlene Bemfica de Oliveira
Instituto Federal de Minas Gerais


Prof(a). Tufi Neder Neto
UFL

Belo Horizonte, 17 de março de 2014.

A todos aqueles que buscam conhecer a realidade dos fatos.

AGRADECIMENTOS

Àquele que me deu vida, saúde e todas as outras condições favoráveis para que eu pudesse realizar esse trabalho. Creio que a força que tive veio Dele.

À minha avó, Anna, à minha mãe, Cornélia, e a toda a minha família, pelo apoio e afeto. À minha mãe por ter me ajudado na preparação para a minha apresentação de Qualificação.

Àqueles que estão mais próximos de mim, Ana Júlia e Luciano, pelo carinho, apoio e paciência.

À Profa. Dra. Heliana Ribeiro de Mello, por sempre acreditar no meu trabalho e me orientar com competência ao longo da realização desta tese.

Ao Prof. Dr. Crysttian Arantes Paixão, por responder meus questionamentos sobre estatística com tanta paciência e pela sua contribuição substancial para a realização das análises estatísticas feitas durante esta pesquisa.

Aos Professores Dr. Ricardo Augusto de Souza, Dra. Deise Prina Dutra, Dra. Shirlene Bemfica, Dr. Tufi Neder Neto e Dr. Guilherme From pelas correções e sugestões para a versão final desta tese.

À Profa. Dra. Deise Prina Dutra, pela sua contribuição para a minha formação acadêmica e profissional.

Aos meus colegas do IFSULDEMINAS, que, nas etapas finais desta tese, me receberam para uma nova etapa da minha vida.

RESUMO

Esta tese trata da variação sintática na expressão de eventos de transferência de posse na interlíngua português/inglês adotando a perspectiva teórico-metodológica da Linguística de *Corpus* (MCENERY & HARDIE, 2012) e partindo dos pressupostos teóricos da Hipótese da Interlíngua (CORDER, 1993; ELLIS & BARKHUIZEN, 2005), da Gramática de Construções (BOAS, 2013; GOLDBERG, 1995) e da Linguística Probabilística (BOD, HAY & JANNEDY, 2003). Os objetivos de pesquisa estabelecidos foram: (I) verificar se brasileiros aprendizes dos níveis intermediário ao avançado de inglês/L2 são sensíveis à associação entre verbos específicos e construções e às propriedades dos argumentos recipiente e tema (acessibilidade discursiva, extensão, pronominalidade e animacidade) que favorecem a seleção de uma das construções dativas do inglês (*Mary gave Paul a book versus Mary gave a book to Paul*); (II) investigar a ocorrência de transferência (ODLIN, 1989) do padrão frasal <Sujeito Verbo Oblíquo_{recipiente} Objeto_{tema}> da L1 dos aprendizes, como em “Maria deu para João um livro”, para suas interlínguas; (III) contribuir com descrições sobre construções dativas do português do Brasil baseadas em dados de corpora. Respostas às perguntas de pesquisa foram buscadas através da análise e comparação de dados extraídos de corpora do português brasileiro (C-ORAL-BRASIL e Lácio-Web), do inglês britânico (ICE-GB) e da interlíngua português/inglês (Br-ICLE e LINDSEI-BR). Para auxiliar a análise dos corpora, o *software* R (GRIES, 2009a) foi utilizado. Os resultados obtidos indicam que: (I) no âmbito de construções da estrutura argumental, brasileiros aprendizes de inglês são capazes não só de adquirir construções da língua-alvo, mas também de selecioná-las em seus contextos preferenciais de uso; (II) a partir do nível de proficiência intermediário na língua-alvo, há pouca probabilidade de transferência para a interlíngua de construções da L1 que apresentam baixa frequência de ocorrência em relação a construções “alternantes”, como é o caso da ordem <Sujeito Verbo Oblíquo_{recipiente} Objeto_{tema}> em relação à ordem <Sujeito Verbo Objeto_{tema} Oblíquo_{recipiente}>; (III) no português do Brasil, as chamadas “alternâncias” nas ordens dos complementos de verbos ditransitivos (Dar uma livro para João *versus* Dar para João um livro) e no uso ou não

das preposições “a” e “para” (Dar para João um livro *versus* Dar a João um livro *versus* Dar João um livro) devem ser caracterizadas em termos das frequências de ocorrência de cada uma das variantes e das diferentes associações entre as variantes e os canais escrito e oral.

Palavras-chave: construções da estrutura argumental, interlíngua, Linguística de *Corpus*.

ABSTRACT

This thesis addresses the syntactic variation in the expression of transfer events by Brazilian learners of English from the theoretical and methodological perspective of *Corpus Linguistics* (MCENERY & HARDIE, 2012). It is also based on the theoretical assumptions of the Interlanguage Hypothesis (CORDER, 1993; ELLIS & BARKHUIZEN, 2005), Construction Grammar (BOAS, 2013; GOLDBERG, 1995), and Probabilistic Linguistics (BOD, HAY & JANNEDY, 2003). The research objectives were to: (I) verify whether Brazilians learners of English at different levels of proficiency (from intermediate to advanced) are sensitive to associations between specific verbs and constructions, as well as to the properties of recipients and themes (discourse accessibility, length, pronominality, animacy) which favor the selection of one of the English dative constructions (Mary gave Paul a book *versus* Mary gave a book to Paul); (II) investigate the transfer (ODLIN, 1989) of the L1 pattern <Subject Verb Oblique_{recipient} Object_{theme}>, as in “Maria deu para João um livro”, to the learners’ interlanguages; (III) contribute to descriptions of Brazilian Portuguese dative constructions based on *corpus* data. Answers to the research questions were sought through the analysis and comparison of data extracted from Brazilian Portuguese corpora (C-ORAL-BRASIL and Lácio-Web), British English corpora (ICE-GB), and Portuguese/English interlanguage corpora (Br-ICLE and LINDSEI-Br). The R software (GRIES, 2009a) was used to aid the corpora analyses. The results of this study indicate that: (I) Brazilians learners of English are able not only to acquire L2 argument structure constructions, but also to select them in their preferred contexts of use; (II) L1 constructions which display low frequency compared to other alternating constructions, as is the case with the Brazilian Portuguese pattern <Subject Verb Oblique_{recipient} Object_{theme}> compared to <Subject Verb Object_{theme} Oblique_{recipient}>, are not likely to be transferred to the learners’ interlanguages after the intermediate level of proficiency in the target-language; (III) in Brazilian Portuguese, the variation in the order of complements of ditransitive verbs (Dar uma livro para João *versus* Dar para João um livro) and in the use or omission of the prepositions “a” and “para” (Dar para João um livro *versus* Dar a João um livro *versus* Dar João um livro) should be characterized in terms of their different frequencies of occurrences and associations with the oral and *Written* channels of

language.

Keywords: argument structure constructions, interlanguage, *Corpus* Linguistics.

LISTA DE FIGURAS, QUADROS E TABELAS

FIGURAS

Figura 1 - Tela do RStudio mostrando janela onde comandos podem ser executados.....	87
Figura 2 - Resultado da execução do comando "tolower(x)" no RStudio.	88
Figura 3 - Resultado da execução do comando "help(gsub)" no RStudio.	89
Figura 4 - Script para a contagem de palavras do subcorpus <i>Written</i> do ICE-GB.....	91
Figura 5 - Parte de um arquivo do <i>corpus</i> ICE-GB.	93
Figura 6 - Reprodução de parte das informações contidas no arquivo gerado com o resultado da execução de um script com a função “grep” para extrair verbos marcados como “ditransitivos” (ditr) no subcorpus <i>Written</i>	94
Figura 7 - Parte de um arquivo do subcorpus <i>Written</i> com destaque para o verbo “render”......	96
Figura 8 - Script do R utilizada durante análise de aspectos semânticos das sentenças.	101
Figura 9 - Script do R utilizado para extrair formas de lemas do C-ORAL-BRASIL.	104
Figura 10 - Parte do resultado do script do R utilizado para a localização de formas do lema [dar] no C-ORAL-BRASIL.	105
Figura 11 - Exemplificação da marcação feita na análise das sentenças do português.	106
Figura 12 - Exemplos de comandos para a execução dos testes qui-quadrado e exato de Fisher no R.....	111
Figura 13 - Resultado da execução de comando com a função prop.test, que utiliza o teste qui-quadrado, no R.....	113
Figura 14 - Resultado da execução de um comando do R utilizado para comparar todas as frequências das categorias semânticas no <i>corpus Written</i> entre si.	114
Figura 15 - Resultados fictícios de uma análise de cluster de consoantes do inglês.	115
Figura 16 - Relação entre os termos presentes na definição de <i>heatmap</i> e as características da análise feita nesta tese.	116
Figura 17 - Exemplo de <i>heatmap</i>	117
Figura 18 - Distribuição das categorias semânticas gerais das expressões por <i>corpus</i>	121
Figura 19 - Gráfico com as distribuições dos padrões de 1 a 5 nos corpora escrito e oral do português analisados.	125
Figura 20 - Gráfico mostrando a distribuição das preposições “a” e “para” nos corpora do português.	129
Figura 21 - Distribuição das construções nos corpora analisados.	130
Figura 22 – Visualização dos p-valores obtidos no R resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes maiores que temas por construção.	157
Figura 23 - Visualização dos p-valores obtidos no R resultantes da aplicação do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes menores que temas por construção.....	157
Figura 24 - Resultado da análise de <i>cluster</i> para comparar as distribuições dos diferentes tipos de recipientes e de temas que ocorrem nas construções ditransitiva e ditransitiva preposicionada nos corpora de inglês L1 e L2.....	159
Figura 25 - Resumo das relações sugeridas entre efeitos de frequência da L1 e da L2 em processos da interlíngua	163

QUADROS

Quadro 1- Exemplos de construções da estrutura argumental	17
Quadro 2 - Proposta de quantidades mínimas de palavras necessárias para se alcançar representatividade em relação a diversos aspectos	30
Quadro 3 - Associações entre padrões sintáticos e estruturas semânticas	42
Quadro 4 - Exemplos de construções de diferentes níveis de representação linguística	43
Quadro 5 - Construções da estrutura argumental do inglês utilizadas na expressão de eventos de transferência de posse	43
Quadro 6- Significados compartilhados entre construções da estrutura argumental e os verbos mais frequentes em cada uma delas	51
Quadro 7 - Variáveis que influenciam a seleção de construções dativas apontadas na literatura.	58
Quadro 8 - Variáveis que favorecem a seleção de cada construção dativa com base na interpretação dos coeficientes obtidos através do teste de Regressão Logística por Bresnan et al. (2007)	61
Quadro 9 - Síntese dos estudos citados nesta seção sobre a alternância dativa do inglês	64
Quadro 10 - Síntese dos estudos citados sobre a alternância dativa na interlíngua	75
Quadro 11 - Síntese dos estudos citados sobre a alternância dativa no português do Brasil	80
Quadro 12 - Resumo das características dos corpora analisados. Indicação dos tamanhos relativos dos corpora baseada na classificação de Sardinha (2000b, p.7)	85
Quadro 13 - Comparação entre as formatações do arquivo do subcorpus <i>Written</i> antes e depois de sua manipulação através de scripts do R	95
Quadro 14 - Exemplos de construções com verbos leves	100
Quadro 15 - Traduções para o português dos verbos do Br-ICLE selecionados para análise. Verbos marcados com * fazem parte da lista dos 100 verbos mais frequentes do C-oral-Brasil.	103
Quadro 16- Número total de expressões selecionadas de cada <i>corpus</i> analisado nesta tese	119

TABELAS

Tabela 1 -Distribuição fictícia de tempos e aspectos do inglês. Fonte: Gries (2003, p. 5)	109
Tabela 2 - Números de verbos cujos contextos foram analisados e números de expressões selecionadas de cada <i>corpus</i> do português	120
Tabela 3 – P-valores resultantes da aplicação do teste qui-quadrado (<i>Humanas</i> e <i>Written</i>) e do teste exato de Fisher (Br-ICLE e LINDSEI-BR) para a comparação das frequências das categorias semânticas por <i>corpus</i>	122
Tabela 4 - Resultados da aplicação do teste exato de Fisher para verificar se as distribuições das categorias semânticas são semelhantes entre os corpora. Destaque para p-valores que levam à aceitação da hipótese nula.	123
Tabela 5 - Distribuição das construções nos corpora do português	124
Tabela 6 - Distribuição das preposições “a” e “para” nos corpora do português	128
Tabela 7 - Distribuição das construções nos corpora de inglês/L1 e L2	130
Tabela 8 - Comparação entre frequências de construções por <i>corpus</i> de inglês L1 e L2 por meio da aplicação do teste exato de Fisher (<i>Written</i> , <i>Spoken</i> e LINDSEI-BR) e do qui-quadrado (Br-ICLE). Destaque para p-valor que leva à aceitação da hipótese nula.	131

Tabela 9 - P-valores obtidos por meio da aplicação do teste exato de Fisher para comparar os corpora de inglês L1 e L2 em relação às distribuições das construções	132
Tabela 10 - Verbos que ocorreram na construção ditransitiva preposicionada invertida no Br-ICLE.....	134
Tabela 11 - Distribuição de verbos por construção no Humanas. Destaque para frequências que indicam com qual construção cada verbo foi mais frequente.	135
Tabela 12 - Distribuição das propriedades dos recipientes e temas nos exemplos da construção ditransitiva preposicionada invertida no Br-ICLE e p-valores resultantes das comparações entre as frequências por coluna por meio do teste exato de Fisher. Destaque para frequências estatisticamente maiores.....	136
Tabela 13 - Distribuição de verbos nas construções ditransitiva (D) e ditransitiva preposicionada (DP) nos corpora de inglês/L1 e L2.....	139
Tabela 14 - P-valores resultantes da aplicação dos testes exato de Fisher e qui-quadrado para a comparação entre as frequências de verbos na construção ditransitiva e na ditransitiva preposicionada por <i>corpus</i> de inglês L1 e L2. Destaque para frequências que indicam com qual construção cada verbo ocorreu mais frequentemente.	141
Tabela 15 - Distribuição de recipientes “pronomes” e “NP lexical” nos corpora de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências de recipientes pronomes e lexicais por <i>corpus</i> de inglês L1 e L2. Destaque para frequências estatisticamente maiores.....	144
Tabela 16 - P-valores resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto à distribuição de recipientes pronomes e NP lexicais	144
Tabela 17 - Distribuição de recipientes pronomes por construção nos corpora de inglês/L1 e L2. Destaque para as frequências de recipientes pronomes na construção ditransitiva.	145
Tabela 18 - Resultado da aplicação do teste exato de Fisher para a comparação entre as frequências de recipientes pronomes que ocorrem na construção ditransitiva nos corpora de inglês L1 e L2	145
Tabela 19 - Frequências de recipientes Np lexicais por construção em cada <i>corpus</i> de inglês L1 e L2 e os p-valores resultantes da aplicação do teste qui-quadrado (<i>Written, Spoken</i> e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação entre essas frequências. Destaque para frequências estatisticamente maiores.	146
Tabela 20 - Distribuição de temas “pronomes” e “NP lexical” nos corpora de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado (<i>Written, Spoken</i> e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação das frequências por coluna. Destaque para frequências estatisticamente maiores.	147
Tabela 21 - Distribuição de recipientes animados e inanimados nos corpora de inglês L1 e L2.....	147
Tabela 22 -Resultados da aplicação do teste qui-quadrado (<i>Written, Spoken</i> e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação entre as distribuições de recipientes animados e inanimados por <i>corpus</i> de inglês L1 e L2.....	148
Tabela 23 - Resultados da aplicação do teste exato de Fisher e do teste qui-quadrado para a comparação das distribuições dos recipientes animados e inanimados nos corpora de inglês L1 e L2	148
Tabela 24 – Frequências de recipiente e temas animados e inanimados por construção nos corpora de inglês L1 e L2 e p-valores resultantes da comparação entre essas frequências por meio da aplicação do teste qui-quadrado e do teste exato de Fisher. Destaque para frequências estatisticamente maiores.....	149
Tabela 25 - Distribuição de recipientes animados por construção nos corpora de inglês/L1 e L2	

.....	150
Tabela 26 - P-valores resultantes da aplicação do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes animados por construção	150
Tabela 27 - Distribuição de construções por recipiente e tema animados e inanimados em cada <i>corpus</i> de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências. Destaque para frequências estatisticamente maiores.	151
Tabela 28 - Resultados da comparação entre as proporções de recipientes dados e novos nos corpora de inglês L1 e L2	152
Tabela 29 - Resultados da comparação entre as proporções de temas dados e novos nos corpora de inglês L1 e L2	152
Tabela 30 - Frequências de construções por recipientes e temas animados e inanimados em cada <i>corpus</i> de inglês/L1 e L2 e p-valores resultantes da comparação entre essas frequências por meio dos testes qui-quadrado e exato de Fisher. Destaque para frequências estatisticamente maiores.	152
Tabela 31 - P-valores resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto à distribuição de recipientes dados na ditransitiva e na ditransitiva preposicionada	153
Tabela 32 - Distribuições de recipientes menores que temas, de recipientes maiores que temas e de recipientes tão extensos quanto os temas por <i>corpus</i> de inglês L1 e L2. Destaque para frequências estatisticamente maiores.	154
Tabela 33 - P-valores resultantes da aplicação do teste qui-quadrado (<i>Written, Spoken, Br-ICLE</i>) e do teste exato de Fisher (LINDSEI-BR) para comparações entre as frequências de recipientes menores que temas, de recipientes maiores que temas e de recipientes tão extensos quanto os temas nos corpora de inglês L1 e L2. Destaque para o p-valor que leva à aceitação da hipótese nula.	154
Tabela 34 – Frequências de recipientes menores, maiores e do mesmo tamanho que temas por construção nos corpora de inglês L1 e L2 e p-valores resultantes da aplicação dos testes qui-quadrado e exato de Fisher para a comparação entre essas frequências. Destaque para frequências estatisticamente maiores.	155
Tabela 35 - P-valores resultantes da aplicação do teste exato de Fisher e do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes menores que temas por construção. Destaque para p-valores que levam à rejeição da hipótese nula.	155

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Contextualização do problema de pesquisa e apresentação dos objetivos	16
1.2 Hipóteses	22
1.3 Importância do estudo	24
1.4 Organização da tese.....	25
2. REVISÃO DA LITERATURA	27
2.1 Linguística de <i>corpus</i>	28
2.2 Abordagem probabilística para estudos linguísticos.....	36
2.3 Abordagem Construcionista	41
2.3.1 Estudos sobre alternâncias entre construções baseados em análise de <i>corpus</i>	47
2.3.2 Aprendendo construções	49
2.4 Interlíngua.....	52
2.4.1 Evidências da existência de construções na Interlíngua.....	55
2.5 Estudos anteriores sobre construções dativas	56
2.5.1 Estudos anteriores sobre construções dativas no inglês	57
2.5.2 Estudos anteriores sobre construções dativas na interlíngua	65
2.5.2.1 Ordem de aquisição das construções dativas	65
2.5.2.2 Padrões sintáticos não esperados na língua-alvo	67
2.5.2.3 Conhecimento sobre quais verbos participam da alternância dativa do inglês	69
2.5.2.4 Sensibilidade de aprendizes aos fatores que influenciam a escolha por uma das construções dativas	72
2.6 Estudos anteriores sobre construções dativas no português do Brasil	75
3. METODOLOGIA.....	82
3.1 Corpora analisados	82
3.2 Software que auxiliou a análise dos corpora	85
3.2.1 Contagem de palavras do subcorpus <i>Written</i> do ICE-GB.....	91

3.2.2 Como o R facilitou a extração das construções dativas do subcorpus <i>Written</i>	94
3.3 Critérios para a seleção de construções dos corpora do inglês e da interlíngua	97
3.4 Definição das categorias para a análise das expressões extraídas dos corpora do inglês e da interlíngua.....	98
3.5 Análise dos corpora do português brasileiro	102
3.6 Análises quantitativas.....	107
3.6.1 Cálculo de frequência.....	107
3.6.2 Testes de significância estatística e análise de cluster	109
 4. RESULTADOS E DISCUSSÃO.....	 118
4.1 Números de expressões selecionadas de cada <i>corpus</i> e caracterização dos universos analisados quanto às categorias semânticas gerais das expressões	118
4.2 Resultados dos corpora do português.....	124
4.3 Distribuição das construções nos corpora de inglês L1 e L2	129
4.4 Construção ditransitiva preposicionada invertida no Br-ICLE	134
4.5 Fatores que influenciam a escolha da construção ditransitiva e ditransitiva preposicionada nos corpora de inglês/L1 e da interlíngua português/inglês	138
4.5.1 Verbo	138
4.5.2 Propriedades dos recipientes e temas	143
 5. CONCLUSÃO.....	 161
 REFERÊNCIAS	 165

1. INTRODUÇÃO

1.1 Contextualização do problema de pesquisa e apresentação dos objetivos

Questões relacionadas à estrutura argumental têm despertado o interesse de diversos pesquisadores (BLEY-VROMAN & JOO, 2001; CUERVO, 2007; GOLDBERG, 1995; PINKER, 1989; GRIMSHAW, 1992; GRIES & STEFANOWITSCH, 2004; INAGAKI, 1997; JOO, 2003; MONTRUL, 1998; OH, 2006; ZARA, OLIVEIRA & SOUZA, 2013). Tradicionalmente, chama-se de estrutura argumental a representação lexical de informações gramaticais sobre um predicator, que, no caso do verbo, especificaria o número de argumentos e seus papéis temáticos, além das posições sintáticas ocupadas pelos mesmos na oração (GRIMSHAW, 1992; PINKER, 1989). Utilizamos o termo “estrutura argumental” nesta tese, contudo, a partir de outra perspectiva, a da Gramática de Construções, para nos referirmos a padrões sintáticos que trazem consigo as interpretações gerais dos eventos expressos pelas orações, ou seja, a estruturas sintáticas abstratas que carregam sentidos que, de certa forma, independem das palavras que nelas ocorrem (BOAS, 2013; CROFT & CRUSE, 2004; GOLDBERG, 1995, 2003; TOMASELLO, 2003a). O Quadro 1 fornece exemplos¹ de construções da estrutura argumental.

¹ Ao longo desta tese, exemplos de frases inventadas são fornecidos quando mencionamos trabalhos anteriores que utilizaram exemplos com frases inventadas. Exemplos de frases extraídas de corpora, por sua vez, são fornecidos quando tratamos dos nossos questionamentos de pesquisa e discutimos nossos resultados.

Quadro 1- Exemplos de construções da estrutura argumental

Construção	Exemplo	Forma	Interpretação geral
Construção transitiva	<i>He sliced the bread.</i>	<Suj. Verb. Obj.>	X age sobre Y
Construção ditransitiva preposicionada	<i>Pat sliced the carrots into the salad.</i>	<Suj. Verb. Obj. Obl.>	X faz com que Y se mova até Z
Construção ditransitiva	<i>Pat sliced Chris a piece of pie.</i>	<Suj. Verb. Obj.1 Obj.2>	X faz com que Y receba Z

Adaptado de Goldberg (2003)

Dentre os fenômenos relativos à estrutura argumental, muita atenção tem sido dada às chamadas “alternâncias da estrutura argumental” (PINKER, 1989), entendidas aqui como a ocorrência do mesmo verbo em diferentes construções da estrutura argumental. Os seguintes pares de sentenças em inglês ilustram algumas dessas alternâncias.

1) Alternância causativa/incoativa (LEVIN, 1993, p. 124):

a) *The strong winds **cleared** the skies.* / Os ventos fortes limparam o céu.

b) *The skies **cleared**.* / O céu clareou.

2) Alternância locativa (BOAS, 2003, p. 1):

a) *Joe **loaded** boxes onto the truck.* / Joe colocou caixas no caminhão.

b) *John **loaded** the truck with boxes.* / Joe carregou o caminhão de caixas.

3) Alternância dativa (BRESNAN et al., 2007, p. 2):

a) *Susan gave toys to the children.* / Susan deu brinquedos para as crianças.

b) *Susan gave the children toys.* / Susan deu para as crianças brinquedos.

A “alternância dativa” do inglês (AD), isto é, a possibilidade de alguns verbos interagirem tanto com a construção ditransitiva preposicionada (3a) quanto com a construção ditransitiva (3b), tem despertado o interesse de muitos pesquisadores devido à observação de que alguns verbos semanticamente muito semelhantes não possuem a mesma aceitação nessas duas construções, como no caso de *give* e *donate*:

4) (a) *I can give details to you.*

(b) I can give you details².

5) (a) Bill Gates donated \$750 million to the fund³.

(b) (?) Bill Gates donated the fund \$750 million.

Instigados por tal diferença entre verbos semanticamente semelhantes, alguns autores tentam encontrar restrições que licenciariam ou não a participação de verbos na alternância dativa (GROEFSEMA, 200; 1PINKER, 1989; RAPPAPORT HOVAV & LEVIN, 2008), enquanto outros questionam se aprendizes de inglês como segunda língua (L2) adquirem conhecimentos sobre essas restrições (INAGAKI, 1997; WHITE, 2003; WHONG-BARR & SCHWARTZ, 2002; ZARA, 2009; ZEDDARI, 2009). Nesta tese, contudo, o interesse pela alternância dativa do inglês surge a partir de outras considerações.

Embora apresentem configurações sintáticas distintas, as sentenças (3a) e (3b) acima são semanticamente semelhantes por compartilham o mesmo verbo e expressarem o mesmo evento

² Ocorrência retirada do subcorpus *Written* do International Corpus of English-GB. Informações sobre o corpus podem ser encontradas em <http://www.ucl.ac.uk/english-usage/projects/ice-gb/>.

³ Ocorrência retirada do texto escrito por Donald G. McNeil Jr., disponível em http://www.nytimes.com/2012/01/27/health/gates-donates-750-million-to-global-fund.html?_r=0.

de transferência no mundo. Isso não quer dizer, porém, que sejam escolhidas aleatoriamente em situações reais de comunicação. De fato, estudos baseados em corpora têm mostrado que construções sintáticas não variam aleatoriamente, antes são escolhidas por falantes sob a influência de fatores linguísticos (ex: associações léxico-gramaticais) e não-linguísticos (ex: canal) (BIBER et al., 1998; KENNEDY, 1998). Surge, assim, um interesse crescente em se investigar quais fatores estão envolvidos na seleção das construções que fazem parte da alternância dativa do inglês como língua materna (inglês/L1) em contextos reais de comunicação (BRESNAN et al., 2007; GRIES, 2003, 2007; GRIES & STEFANOWITSCH, 2004; KENDALL et al., 2011; THEIJSSSE, 2008). A partir de resultados de estudos dessa agenda de pesquisa, questiona-se se aprendizes de inglês/L2 são sensíveis aos fatores que favorecem o uso de cada construção dativa do inglês (CALLIES & SZCZESNIAK, 2008). É esse mesmo questionamento que colocamos nesta tese investigando o uso dessas construções por aprendizes brasileiros.

Ao definirmos como objeto de estudo produções de falantes de uma segunda língua (L2), partimos do pressuposto de que, na aquisição de uma L2, falantes desenvolvem, com base no *input* da L2 e influenciados por diversos fatores, um sistema linguístico internamente estruturado, que possui elementos da L1, da L2 e elementos ausentes nessas duas línguas, ao qual se dá o nome de Interlíngua (BARRON, 2003; CORDER, 1993; ELLIS, 1994; ELLIS & BARKHUIZEN, 2005; KRAMSCH & WHITESIDE, 2007; NEMSER, 1971; SELINKER, 1972). Dentre os fatores que influenciam o desenvolvimento da Interlíngua, muita atenção tem sido dada à língua materna dos aprendizes (BARRON, 2003; BROOKE & HIRST, 2012; CORDER, 1993; ELLIS, 1985; JARVIS, 2000; KELLERMAN, 1979; LAKSHMANAN & SELINKER, 2001; ODLIN, 1989; OSBORNE, 2008; SCHACHTER, 1993; SELINKER, 1972), já que o uso de conhecimentos prévios em novas situações de aprendizagem é um processo psicológico geral (ELLIS, 1985) e que produções da Interlíngua que se aproximam de estruturas da L1, mas não de estruturas da língua-alvo, são documentadas pela literatura (BROWN, 2007). Compartilhando esse mesmo interesse, e assumindo que adquirir construções da estrutura argumental corresponde a aprender como expressar linguisticamente cenas básicas da experiência humana (GOLDBERG, 1995), esta tese também tem como objetivo investigar como o conhecimento de português/L1 influencia a interlíngua português/inglês no que tange à expressão linguística de eventos de transferência de posse.

Evidências que apontam para a influência do português na interlíngua português/inglês no que diz respeito a aspectos da estrutura argumental são fornecidas por estudos anteriores (SOUZA, 2011; VILELA, 2009; ZARA, 2009). Focalizando as construções dativas do inglês, Zara (Ibidem) relata resultados que apontam para a influência da L1 na ordem de aquisição dessas construções. Através de uma tarefa de julgamento de aceitabilidade de sentenças, a pesquisadora constatou que sentenças em inglês com a construção ditransitiva preposicionada, presente no português do Brasil (PB), foram amplamente aceitas por falantes da interlíngua português/inglês desde o nível iniciante de proficiência em inglês/L2. Por outro lado, ela aponta que sentenças com a construção ditransitiva, documentada apenas em algumas variedades do PB, foram amplamente aceitas apenas no nível avançado de proficiência em inglês/L2. Além disso, Zara observou que sentenças em inglês como **Mary gave to John a letter*, que apresentam uma estrutura presente na L1 dos sujeitos da sua pesquisa, <Sujeito Verbo Oblíquo Objeto>, foram mais aceitas pelos aprendizes do nível iniciante de proficiência na língua-alvo do que sentenças como *Mary gave John a letter*. Como conclusão de seu trabalho, a autora sugere que a influência da L1 na interlíngua pode explicar a ordem de aquisição das construções dativas do inglês por aprendizes brasileiros, que é oposta àquela observada na aquisição de inglês como língua materna (CAMPBELL & TOMASELLO, 2001), e o fato de aprendizes brasileiros no estágio inicial de aquisição de inglês/L2 preferirem sentenças com um padrão sintático que não é o esperado na língua alvo, porém que corresponde a uma tradução direta de um padrão de sua L1 (*Maria deu para João um livro* → *Mary gave to John a book*). Tais conclusões, baseadas em julgamentos de aceitabilidade de sentenças, motivam-nos a investigar a influência da L1 a partir de dados de produção da interlíngua português/inglês.

Discussões sobre a influência da língua materna na interlíngua requerem descrições adequadas dos aspectos da L1 considerados passíveis de transferência. Partindo da hipótese de que a transferência linguística, isto é, o uso de elementos da L1 na performance em L2 (SELINKER, 1972), é influenciada, dentre outros, pela frequência de construções da L1 (POULISSE & BONGAERTS, 1994), a descrição da L1 sobre a qual falamos, para ser mais completa, deve incluir dados de frequência dos aspectos da L1 cuja influência sobre a interlíngua se quer investigar.

Com base nos pressupostos e questionamentos supracitados, três objetivos gerais foram estabelecidos para esta tese:

- (I) Fazer uma descrição da interlíngua português/inglês a partir do nível intermediário de proficiência na L2 no que tange ao uso das construções dativas do inglês.
- (II) Investigar a influência do português na expressão de eventos de transferência de posse por brasileiros aprendizes de inglês/L2.
- (III) Contribuir para descrições sobre construções dativas do português do Brasil com dados de frequência observados em corpora.

Esses objetivos gerais se desdobram nos seguintes objetivos específicos:

- (I) Identificar quais construções são mais usadas para expressar eventos de transferência de posse quando o recipiente e o tema são linguisticamente expressos como complementos dos verbos no português, no inglês/L1 e na interlíngua português/inglês.
- (II) Verificar a associação entre os fatores abaixo e as construções dativas do português/L1 (fatores a e b) e do inglês L1 e L2 (fatores a-c):
 - a) Canal: escrito e oral.
 - b) Verbo utilizado.
 - c) Propriedades dos argumentos <recipiente> e <tema> (definidas na seção 3.4 desta tese): pronominalidade, complexidade sintática, acessibilidade discursiva e animacidade.
- (III) Descrever amostras de variedades do português do Brasil escrito e falado quanto às frequências:
 - a) Dos diferentes padrões frasais utilizados para expressar eventos de transferência de posse: (1) <Verbo Objeto_{Tema} Oblíquo_{Recipiente}>, “dar a bola para a garotada” (exemplo do *subcorpus* Humanas); (2) <Verbo Oblíquo_{Recipiente} Objeto_{Tema}>, “disseram a ela toda a verdade” (exemplo do *subcorpus* Humanas); (3) <Verbo Objeto_{Recipiente} Objeto_{Tema}>, “Ele nu dá ocê a mão não” (exemplo do *corpus* C-ORAL-BRASIL); (4)

<Pronome_{Recipiente} Verbo Objeto_{Tema} >, “nenhum dirigente nos deu esse comando” (exemplo do subcorpus Humanas); (5) < Verbo Objeto_{Tema} Objeto_{Recipiente}>, “Deus deu dinheiro ocês” (exemplo do *corpus* C-ORAL-BRASIL).

- b) Do uso das preposições “a” e “para” nos padrões <Sujeito Verbo Objeto Oblíquo> e <Sujeito Verbo Oblíquo Objeto>.

(IV) Identificar em quais aspectos as produções da interlíngua português/inglês que expressam eventos de transferência de posse aproximam-se do português, do inglês e/ou diferenciam-se destas duas línguas.

1.2 Hipóteses

Sobre os dados linguísticos a serem analisados, levantamos as seguintes hipóteses com base em resultados de estudos anteriores, como explicitamos para cada hipótese:

(I) Os brasileiros aprendizes dos níveis intermediário ao avançado de proficiência na L2 devem ser sensíveis às associações entre os fatores supracitados e as construções dativas do inglês. Estudos anteriores já constataram que aprendizes de inglês falantes nativos de outras línguas demonstram-se sensíveis a algumas dessas associações (CALLIES & SZCZESNIAK, 2008; GRIES & WULFF, 2005).

(II) Alguma diferença entre os corpora escritos e orais quanto a aspectos das construções investigadas deve ser identificada, já que o canal é um fator que exerce influência sobre o uso de construções linguísticas (GRIES & STEFANOWITSCH, 2008; MCENERY & HARDIE, 2012; RASO, 2012).

(III) Nos dados do português, a ordem <Sujeito Verbo Objeto Oblíquo> deve ser a mais

frequente, a omissão da preposição, como em *Dei Maria um livro* ou *Dei livro Maria*, deve ser rara, e a preposição “para” deve ser mais frequente do que a preposição “a” na fala, conforme aponta Gomes (2003).

(IV) A transferência (ODLIN, 1989) do padrão sintático <Sujeito Verbo Oblíquo Objeto> do português para a interlíngua pode ocorrer. Zara (2009) fornece os exemplos abaixo dessa transferência em produções espontâneas da interlíngua português/inglês coletadas através de observações informais não-sistemáticas durante sua atuação como professora de inglês.

- a) I am **sending to you** the writings of units 1 and 2 (produção escrita).
- b) They came to this school and **gave to the students** a camera to take pictures of the school (produção escrita).
- c) I bought this book and the Harlequin Books Publishing House **sent to me** a new book (produção escrita).
- d) The soul who has made a lot of adventures in different places, **has offered to him** the wisdom and money (produção escrita).
- e) Please, **bring to us** the bill (produção oral).
- f) Do you intend to **give to Ana Júlia** the same opportunities? (produção oral).

Porém, com base nos dados quantitativos obtidos pela mesma pesquisadora, espera-se que a ocorrência desse tipo de transferência seja baixa.

1.3 Importância do estudo

Desde a década de 1960, a Interlíngua tem sido o foco de muitos estudos empíricos que se preocupam em identificar a natureza e o desenvolvimento do conhecimento sobre a L2 (GASS & SELINKER, 1994). Inseridas nesse contexto, muitas pesquisas têm focalizado questões acerca da aquisição da estrutura argumental da L2 (BLEY-VROMAN & JOO, 2001; INAGAKI, 1997; JOO, 2003; WHONG-BARR & SCHWARTZ, 2002; WHITE, 2003; ZARA, 2009; ZEDDARI, 2009). No entanto, tais investigações têm geralmente se limitado a analisar a aquisição de construções inexistentes na L1 dos aprendizes ou o sucesso na categorização dos verbos que podem ou não interagir com construções da L2, ou seja, o aprendizado sobre quais verbos participam das chamadas “alternâncias da estrutura argumental”. No que tange à alternância dativa em específico, poucos são os estudos que investigam quais aspectos influenciam a alternância na linguagem do aprendiz (CALLIES & SZCZESNIAK, 2008, p. 169). Portanto, nossa proposta de investigar o conhecimento de aprendizes sobre os contextos preferenciais de uso de construções da estrutura argumental da L2 constitui um tipo de estudo relevante que vai além de questões tradicionalmente exploradas. Nossa proposta parte também da consideração de que proficiência em uma língua envolve não apenas o conhecimento das construções que dela fazem parte, mas também a capacidade de selecionar essas construções em seus contextos preferenciais de uso (WULFF & GRIES, 2006).

A reconhecida importância da L1 no processo de aquisição da L2 (BENSON, 2002; BLEY-VROMAN & JOO, 2001; CORDER, 1993; ELLIS & BARKHUIZEN, 2005; HOPP, 2010; JARVIS, 2000, 2011; ODLIN, 1989) justifica nossa preocupação em investigar a influência do português na aquisição do inglês por aprendizes brasileiros. No caso específico da aquisição da estrutura argumental da L2, evidências empíricas fornecem dados que apontam para a influência da L1 nesse processo (INAGAKI, 1997; OH, 2006; SOUZA, 2011; WHITE, 2003). Esperamos, assim, que os dados da interlíngua português/inglês analisados nesta tese fomentem discussões teóricas acerca da influência translinguística.

Nossa investigação sobre a Interlíngua, contextualizada acima, é baseada em análise de corpora (MCENERY & WILSON, 2001). Essa opção metodológica é outra contribuição desta

pesquisa, pois a linguagem produzida pelos aprendizes é fonte central de evidência para se ter acesso às representações da Interlíngua (MYLES, 2005). Além disso, na busca pela compreensão da Interlíngua, exploraremos também a língua materna dos aprendizes através da análise de corpora de variedades do português do Brasil. Assim, ao olharmos para a influência da L1 na Interlíngua a partir da abordagem proposta, contribuímos com descrições sobre a L1 baseadas em dados empíricos mais generalizáveis.

Por fim, destacamos que as análises de *corpus* feitas nesta pesquisa foram auxiliadas pelo software R. Esperamos que a apresentação que fazemos dos recursos disponíveis desse software no capítulo sobre a metodologia por nós utilizada seja um convite para que mais linguistas no Brasil possam se favorecer das vantagens que o R oferece, assim como linguistas de outras partes do mundo já têm feito (BAAYEN, 2008; GRIES, 2009a).

1.4 Organização da tese

Esta tese foi organizada em 5 capítulos. Neste capítulo introdutório, contextualizamos nosso estudo no cenário de questionamentos já feitos sobre a estrutura argumental, apresentamos nossos objetivos de pesquisa e hipóteses, e destacamos a importância deste trabalho. No capítulo de revisão da literatura, capítulo 2, apresentamos as bases teórico-metodológicas nas quais nos fundamentamos: a Linguística de *Corpus*, a Linguística Probabilística, a Gramática de Construções e a Hipótese da Interlíngua. Além disso, neste mesmo capítulo fazemos uma revisão de alguns estudos sobre construções dativas em algumas variedades do inglês, em interlínguas de falantes nativos de diversas línguas e em algumas variedades do português do Brasil. É a partir dos resultados desses estudos que levantamos nossas hipóteses sobre a interlíngua português/inglês.

No capítulo sobre a metodologia utilizada para alcançarmos nossos objetivos, capítulo 3,

descrevemos primeiramente os corpora analisados. Em seguida, apresentamos de forma breve o *software* que nos auxiliou na análise dos corpora e explicamos como ele foi utilizado nesta pesquisa. Na sequência, explicitamos quais critérios foram utilizados para a seleção de expressões dos corpora, definimos as categorias para a análise dos corpora do inglês e da interlíngua e descrevemos, separadamente, como se deu a análise dos corpora do português. Finalmente, mostramos como foram feitas as análises quantitativas dos dados. Relatamos e discutimos os resultados obtidos no capítulo 4. Apresentamos as conclusões do estudo, assim como prospecções de pesquisas futuras, no capítulo 5.

2. REVISÃO DA LITERATURA

Nosso estudo se insere na agenda de pesquisa que investiga como as pessoas exploram os recursos de suas línguas para expressarem significados, buscando motivações para a existência de múltiplas estruturas na língua que são semelhantes quanto à forma e/ou sentido. Nessa tarefa, a Linguística de *Corpus* (BIBER et al., 1998; MCENERY & HARDIE, 2012; MCENERY & WILSON, 2001) tem se mostrado promissora, e, por essa razão, optamos por adotá-la como modelo teórico-metodológico.

A Linguística de *Corpus* não faz afirmações absolutas sobre o que sempre ocorre ou nunca ocorre na língua, antes admite que os padrões linguísticos ocorrem em proporções variadas que devem ser indicadas por medidas quantitativas (BIBER et al., 1998). Nesse sentido, a Linguística de *Corpus* reconhece a natureza probabilística da língua, como propõe a Linguística Probabilística (BOD, HAY & JANNEDY, 2003). Ambos os modelos adotam uma perspectiva para os estudos linguísticos que se opõe a abordagens categoriais, considerando a variabilidade no uso da língua como um aspecto a ser incluído na descrição gramatical. É essa a posição que defendemos também nesta tese.

A observação da língua em contextos específicos de uso revela que padrões linguísticos são associados a funções semânticas, discursivas e apresentam outras particularidades quanto ao seu uso. Uma proposta para considerar esses tipos de informações como parte da gramática é encontrada em abordagens construcionistas (CROFT & CRUSE, 2004; ELLIS, 2003; ELLIS & LARSEN-FREEMAN, 2009; GOLDBERG, 1995; LANGACKER, 1987; PALMER, 1996; TOMASELLO, 2006). Baseamo-nos em tais abordagens para complementarmos nossa compreensão sobre a natureza e a aquisição da língua. Ao analisarmos a natureza e a aquisição da L2 em particular, partimos da Hipótese da Interlíngua (CORDER, 1993; ELLIS, 1997; SELINKER, 1972).

As características de cada um dos modelos teóricos supracitados que consideramos mais relevantes dentro do escopo desta tese são discutidas nas primeiras 4 seções deste capítulo. Em seguida, são revisados alguns trabalhos sobre construções dativas em variedades do inglês, da

interlíngua e do português, a partir dos quais nossas perguntas de pesquisa surgiram e nossas hipóteses foram formuladas.

2.1 Linguística de *corpus*

Seguimos Arppe et al. (2010) e defendemos que o uso efetivo da língua em sua variedade e riqueza deve ser explorado para que se chegue a hipóteses mais confiáveis e completas sobre fenômenos linguísticos. Essa orientação teórico-metodológica está em consonância com a Linguística de *Corpus* (LC), que se encarrega da coleta e exploração de corpora com a finalidade de estudar a língua a partir de exemplos de seu uso real (BIBER et al., 1998; KENNEDY, 1998; MCENERY et al., 2006; MCENERY & HARDIE, 2012; MCENERY & WILSON, 2001; SARDINHA, 2000b).

A concepção atual de *corpus* na LC corresponde a um conjunto estruturado de textos escritos ou orais produzidos em situações reais de comunicação, eletronicamente armazenados e processados, que foram coletados de acordo com critérios específicos (conteúdo, gênero, tipologia, registro etc.) para fins de análise linguística (SARDINHA, 2004). Essa definição aponta para quatro características fundamentais de um *corpus* (MCENERY & HARDIE, 2012; MCENERY & WILSON, 2001):

- (I) Formato eletrônico: os textos que compõem um *corpus* devem estar no formato eletrônico para que o *corpus* seja manipulado e pesquisado de forma mais fácil e rápida através do computador.
- (II) Tamanho finito: um *corpus* geralmente tem um tamanho finito (ex: 1 milhão de palavras), com exceção de corpora monitores, aos quais textos são constantemente adicionados. Consideramos relevante lembrar que, seja qual for o tipo de *corpus*, um *corpus* é necessariamente um subconjunto finito de conjunto muito maior, em

princípio infinito, que é a língua (MECENERY & HARDIE, 2012).

- (III) Função representativa: entende-se que um *corpus* é uma referência padrão para certa variedade de língua, devendo conter uma amostragem suficiente de uma variedade da língua para ser representativo desta. Devido à característica (II) acima, alguns autores consideram difícil alcançar o ideal de “representatividade” (Ibidem).

A representatividade de um *corpus* tem sido normalmente relacionada com o seu tamanho. A justificativa para se relacionar a extensão de um *corpus* com a sua representatividade fundamenta-se no reconhecimento da natureza probabilística da língua. Como explica Sardinha (2000b, p. 5), “algumas palavras têm frequência de ocorrência muito rara, e para que haja probabilidade de elas ocorrerem no *corpus* é necessário incorporar-se uma quantidade grande de palavras ao *corpus*”. Similarmente, no tocante aos sentidos das palavras, o autor afirma que para cada item lexical existem sentidos mais frequentes e menos frequentes. Consequentemente, “mesmo palavras de alta frequência tem sentidos raros (por exemplo, ‘serviço’ entendido como ‘saque’ no jogo de tênis), e, portanto, esses sentidos terão maior probabilidade de ocorrerem quanto maior for o *corpus*” (Ibidem).

Por falta de medidas claras em relação à pergunta “Que número de dados seria suficiente para que um *corpus* seja representativo?”, a posição geralmente adotada, proposta originalmente por Sinclair (1991 *apud* SARDINHA, 2000b, p. 2), é a de que “quanto maior, melhor”. Trabalhar com amostras de tamanhos mais próximos dos “ideais”, porém, nem sempre é possível devido a restrições de tempo e de outros recursos necessários para a compilação de um *corpus*. É relevante saber, portanto, qual seria “o mínimo” necessário para se ter um *corpus* representativo. Nesse intuito, Biber (1990, 1993 *apud* SARDINHA, 2000b) propõe uma abordagem estatística para definir o tamanho mínimo de uma amostragem em termos de número total de textos, número de textos por variedade e tamanho de texto para que ela seja representativa de certa variedade linguística, conservando suas características estáveis. A metodologia utilizada pelo autor parte da comparação entre diversos traços linguísticos presentes em uma amostra maior e em amostras menores de um mesmo *corpus*. Em relação ao número de textos, os resultados de Biber mostraram que o número mínimo varia de acordo com o traço linguístico considerado: traços mais frequentes exigem uma quantidade menor e traços menos frequentes exigem uma quantidade maior de textos.

Baseado nos trabalhos de Biber, Sardinha (2000b) se propôs a calcular qual o tamanho mínimo de uma amostra em termos de números de palavras para que a mesma represente estatisticamente certos traços linguísticos. O autor aplicou a mesma metodologia de estimação de amostras baseada em cálculos estatísticos de Biber em 5 corpora: 1 do inglês americano (Brown), 3 do inglês britânico (LOB, BNC-S, “Cartas de pedido de emprego”) e 1 do português brasileiro (Ensaio). Os resultados alcançados pelo autor levaram aos números mostrados no Quadro 2, relativos às quantidades mínimas de palavras necessárias para se alcançar representatividade em relação a diversos aspectos linguísticos:

Quadro 2 - Proposta de quantidades mínimas de palavras necessárias para se alcançar representatividade em relação a diversos aspectos

Tipo de representatividade	Mínimo de palavras
De uma língua como um todo	190 mil
De um <i>corpus</i> especializado	91 mil
Da categoria “verbos”	67 mil
Da categoria “substantivos”	75 mil
De “adjetivos”	150 mil
De “advérbios”	206 mil
Das “palavras de conteúdo”	206 mil

Adaptado de Sardinha (2000b)

Além do tamanho em termos de número de palavras, outros aspectos de um *corpus* influenciam a sua representatividade. Do ponto de vista da Estatística, a representatividade é alcançada quando uma amostra contém em proporção tudo o que a população possui qualitativa e quantitativamente (COSTA, 1992, p. 27). Essa mesma noção aplicada a um *corpus* corresponde a dizer que um *corpus* é representativo na medida em que ele inclui a gama de variabilidade observada em uma população (BIBER, 1993, p. 243). Em corpora compilados para serem representativos de uma língua como um todo, essa variabilidade é alcançada quando dele fazem parte gêneros, registros e tipos textuais variados. Já em corpora de tipos textuais, gêneros ou registros específicos, um maior número de textos distintos representando diferentes autores é necessário para garantir a variabilidade sobre a qual Biber fala.

Cada *corpus* deve ser, então, cuidadosamente desenhado para ser representativo de uma língua ou variedade linguística. As informações que um *corpus* pode trazer sobre uma variedade linguística são diversas, havendo a necessidade de classificar os corpora de acordo com os

seguintes critérios (SARDINHA, 2004):

a) Autor:

Corpus de língua nativa: os autores dos textos são falantes nativos da língua representada pelo *corpus*.

Corpus de aprendiz: os autores dos textos não são falantes nativos da língua representada pelo *corpus*.

b) Língua:

Corpus monolíngue: inclui textos escritos em apenas uma língua ou variedade linguística.

Corpus multilíngue: inclui idiomas distintos.

c) Modo:

Corpus falado: composto por arquivos de som, transcrições ortográficas e/ou fonéticas de monólogos, diálogos e conversas, e anotações prosódicas.

Corpus escrito: composto por textos escritos.

d) Tempo:

Corpus sincrônico: representa uma variedade linguística em um período de tempo.

Corpus diacrônico: representa uma variedade linguística incluindo vários períodos de tempo.

Corpus contemporâneo: representa o período de tempo atual.

Corpus histórico: representa um período de tempo passado.

e) Seleção:

Corpus estático: representa uma amostra finita de uma variedade linguística.

Corpus monitor: sua composição é constantemente modificada de forma a refletir o estado atual de uma variedade linguística.

f) Conteúdo:

Corpus representativo da língua geral: composto por um número balanceado de tipos de textos, gêneros e registros.

Corpus especializado: composto por textos de tipos, gêneros ou registros específicos.

Corpus regional ou dialetal: se composto por textos de uma ou mais variedades sociolinguísticas específicas.

g) Disposição interna:

Corpus comparável: composto por dois ou mais subcorpora com textos escritos originalmente nas línguas ou variedades linguísticas representadas nos subcorpora e que compartilhem alguma característica que seja tomada como referência para a comparação entre os subcorpora (gênero, tipologia textual, data de publicação etc.).

Corpus paralelo: apresenta textos que são comparáveis (ex: original e tradução).

Corpus alinhado: apresenta traduções abaixo das linhas originais.

h) Nível de codificação:

Corpus não- anotado: formado apenas por textos “puros” (raw text) como retirados de suas fontes.

Corpus anotado: enriquecido com diversas informações linguísticas (ex: informações morfossintáticas) e não-linguísticas (ex: registro, idade e sexo do falante/escritor) sobre os textos.

Podemos acrescentar à lista acima os relativamente novos corpora em vídeo, que registram traços como gestos, e os corpora de línguas de sinais (MCENERY & HARDIE, 2012, p. 3).

As finalidades dos corpora são muitas. Santos (2008, p. 51) as classifica em quatro tipos: (I) “ter uma ideia do problema/conhecer” a realidade linguística, (II) “medir um dado fenômeno” linguístico, (III) avaliar “uma hipótese, um sistema, um método, uma teoria”, e (IV) “criar outras coisas”, como dicionários, materiais de ensino de línguas, sistemas de resposta automática e perguntas, sistemas de detecção de plágio etc. Gries (2009b, p. 5), por sua vez, destaca que dados

de frequência obtidos através de análise de corpora (frequências de ocorrência de elementos linguísticos e frequências de co-ocorrência desses elementos) podem ser usados para investigar uma série de questões envolvendo a aquisição de língua materna, a aquisição de uma L2, a relação entre língua e cultura, desenvolvimentos históricos das línguas, além de aspectos fonológicos, morfológicos, sintáticos, semânticos e pragmáticos. Exemplos de perguntas de pesquisa incluem “Há relação entre a frequência de padrões no insumo da língua e a facilidade com que crianças aprendem esses padrões de sua L1?”, “Diferenças entre frequências de palavras refletem diferenças entre culturas?” e “Como prever as escolhas sintáticas de falantes/escritores?” (Ibidem).

Mas por que recorrer a um *corpus* se temos fácil acesso às intuições de falantes proficientes sobre a língua? Afinal, muitas descrições que fazem parte da teoria linguística influenciada pela publicação de *Aspects of the Theory of Syntax* por Chomsky em 1965 baseiam-se na introspecção de falantes, como lembra Teubert (2007, p. 50). Contudo, a observação de que algumas “regras” presentes em tais descrições são “violadas” por textos e a alegação de que outras “regras” não descrevem de forma suficientemente satisfatória alguns aspectos da língua, levaram alguns linguistas à conclusão de que dados reais de uso da língua são necessários para se descrever adequadamente a mesma (Ibidem, p. 51). Vejamos porque aqueles linguistas tinham razão.

Primeiramente, já foi observado que a forma como as pessoas afirmam usar a língua se difere muito da forma como elas de fato a usam (MANNING, 2003, p. 292), o que impossibilita uma descrição linguística mais próxima da realidade a partir da percepção dos falantes. Além disso, sabe-se que ocorrências atípicas tendem a ser mais notadas por falantes do que ocorrências típicas (BIBER et al., 1998), o que pode levar a descrições distorcidas sobre a língua. É preciso considerar também que descrições baseadas na intuição ou em dados de poucos falantes podem revelar nada mais que idiossincrasias que não possuem valor representativo (Ibidem) ou aspectos de variedades linguísticas específicas (XIAO, 2009, p. 989). Por fim, dados introspectivos são descontextualizados, e o contexto é particularmente relevante para julgamentos de aceitabilidade e gramaticalidade (Ibidem) e é fundamental para a construção de significados (SINCLAIR, 2004).

Existe ainda outro argumento a favor de estudos linguísticos baseados em corpora. Se a

Linguística se interessa por generalizações e explicações sobre diversos fenômenos linguísticos que fazem parte da competência linguística, mas essa competência não pode ser diretamente observada, é necessário, então, buscar evidências dessa competência em seu reflexo na língua usada por falantes/escritores (BARONI & EVERT, 2009, p. 777). Areladas a esse argumento estão a concepção de língua como o conjunto de expressões produzidas por falantes/escritores e o pressuposto de que evidências obtidas através da análise do uso da língua podem e devem ser a base para a construção de teorias sobre o conhecimento linguístico (XIAO, 2009, p. 989). Parte-se da premissa maior de que qualquer teoria que não dá conta de dados autênticos é falsa (Ibidem, p. 988).

No caso específico desta tese, a escolha pela LC como base teórico-metodológica justifica-se também pela afinidade entre nossas perguntas de pesquisa e os dois objetivos apontados por Biber e colegas (1998, p. 3) como centrais em pesquisas baseadas em corpora. O primeiro deles é avaliar em que medida um padrão é encontrado em uma língua e o segundo é analisar os fatores contextuais que influenciam a variabilidade linguística. No tocante ao primeiro objetivo, ao invés da postulação de julgamentos categóricos de gramaticalidade, padrões típicos de uso são apontados, como pretendemos fazer com a investigação de quais padrões são usados para a expressão de eventos de transferência de posse na interlíngua português/inglês. No que diz respeito ao segundo objetivo, a LC preocupa-se em explicar como estruturas aparentemente similares ocorrem em contextos diferentes, o que também propomos fazer aos investigarmos a escolha por uma das construções dativas do inglês for falantes da interlíngua português/inglês.

Como mostra a discussão acima, para descrever os padrões de uso da língua e explicar sua variabilidade, a LC vai além de descrições gramaticais tradicionais. A chave para isso é a busca por associações sistemáticas entre estruturas linguísticas (associações lexicais e associações gramaticais), e por associações entre estruturas linguísticas e fatores não linguísticos, como diferentes registros, períodos de tempo e sexo dos falantes (BIBER et al., 1998, p. 6). Um exemplo de como associações lexicais podem levar a descrições mais apuradas sobre as palavras é encontrado na diferenciação entre os adjetivos do inglês *big*, *large* e *great* feita a partir da análise da co-ocorrência desses adjetivos com diferentes grupos de substantivos feita por Biber e colegas (Ibidem, p. 45). A análise dos 10 colocados mais frequentes com cada um desses adjetivos levou os autores às seguintes conclusões: (I) “*big*” é mais comum para se referir a

“tamanho físico” (*big house*), (II) “*large*” é mais frequentemente usado para se referir a “quantidade de algo” (*large number of people*), (III) “*great*”, além de ser muito usado para mostrar “quantidade”, apresenta uma gama maior se significados possíveis, como “intensidade” (*great importance*) e sentido de “muito bom” (*take great care of him*).

Associações entre estruturas gramaticais e outros fatores linguísticos e não-linguísticos, por sua vez, são alvos de estudos baseados em corpora graças à disponibilização de corpora de tamanhos razoáveis contendo anotações gramaticais e diversas informações sobre os textos, e ao desenvolvimento de novos métodos computacionais. Nesselhauf e Romer (2007), por exemplo, fizeram uma análise de *corpus* para identificar quais fatores influenciam o uso de formas progressivas com referência ao futuro em inglês.

Padrões de associação como os supracitados representam relações quantitativas. Por exemplo, pode-se calcular com que frequência morfemas, palavras e padrões gramaticais são encontrados em um *corpus*, com que frequência morfemas são usados com palavras particulares, com que frequência palavras são ocorrem em certos padrões gramaticais etc. (GRIES, 2009a). Embora alguns padrões possam ser identificados diretamente a partir da observação de dados de frequência, é preciso considerar que qualquer amostragem de uma variedade linguística é sujeita à variação aleatória, isto é, à variação fruto do acaso (BARONI & EVERT, 2009, p. 778). Por isso, testes de significância estatística são necessários em alguns casos. Esses testes mostram com que probabilidade as diferenças quantitativas observadas são motivadas ou não. Análises estatísticas ainda mais elaboradas são necessárias para se investigar complexos padrões de associação. Portanto, o linguista de *corpus* deve adquirir conhecimentos sobre estatística ou trabalhar juntamente com um estatístico que o ajude a responder algumas perguntas de pesquisa.

Ao olhar para dados estatísticos sobre a língua, o linguista deve ter em mente uma propriedade estatística da língua observada em análises de corpora chamada de “Lei de Zipf”. Santos (2008, p. 55) explica que essa lei expressa o fato de “haver sempre muitos casos diferentes com pouca frequência, enquanto que os casos de alta frequência correspondem sempre a poucos casos”. Em outras palavras, na análise da distribuição quantitativa de estruturas linguísticas e de padrões de associação, observa-se que existem mais casos diferentes com menor frequência, de forma que padrões menos frequentes se tornam quase a maioria da língua. Assim, se eliminarmos esses casos, ficaremos apenas com uma pequena porção da língua para ser analisada, e, conseqüentemente, teremos uma descrição bastante falha da mesma. Por essa razão, é relevante

considerarmos a totalidade dos dados linguísticos.

Destacamos acima a importância da análise quantitativa de dados de corpora. Ressaltamos, porém, que o objetivo da LC não é apenas relatar dados quantitativos, mas sim explorar a importância desses dados para se compreender os padrões de uso da língua (BIBER et al., 1998, p. 9). Assim, para a LC, tanto abordagens qualitativas quanto abordagens quantitativas para a análise de dados são necessárias (MCENERY & WILSON, 2001). Enquanto a análise quantitativa fornece resultados estatisticamente confiáveis e generalizáveis, a análise qualitativa de um *corpus* favorece uma interpretação desses resultados e uma descrição rica e precisa dos fenômenos linguísticos, sendo também base para a formação de categorias para mais análises.

Resumindo a discussão feita até agora nesta seção, podemos dizer que a LC traz uma série de vantagens para estudos linguísticos. Com a disponibilidade de corpora eletrônicos, é possível (I) pesquisar, recuperar, organizar e realizar cálculos com um número de dados linguísticos antes inimaginável, (II) analisar a língua de forma mais objetiva e sistemática, em oposição a métodos introspectivos, (III) controlar diversos fatores contextuais ao mesmo tempo e (IV) ter acesso a diversas informações sobre os contextos de produção dos textos, tais como a idade e o sexo do autor, a variedade da língua utilizada, o gênero textual etc. Com tantas vantagens, abordagens baseadas em corpora são atualmente utilizadas em praticamente todos os ramos da Linguística (XIAO, 2009, p. 991).

Apontamos mais acima que uma das razões para linguistas voltarem sua atenção para o uso efetivo da língua é o fato de existirem dados empíricos que vão de encontro ao que assumem alguns modelos gramaticais categoriais baseados na introspecção e/ou em sentenças produzidas por poucos falantes. Voltamos a abordar essa questão na seção 2.2, quando introduzimos outro modelo teórico que compartilha características com a Linguística de *Corpus*: a Linguística Probabilística.

2.2 Abordagem probabilística para estudos linguísticos

Vimos na introdução desta tese que um mesmo verbo pode ser encontrado em diferentes construções da estrutura argumental. As sentenças abaixo (KEMMERER, 2006, p. 350) fornecem

mais exemplos desse fenômeno.

6) Ocorrência do verbo *kick* (chutar) em diferentes padrões sintáticos:

a) Bill kicked the ball. / Bill chutou a bola.

b) Bill kicked the ball into the lake. / Bill chutou a bola no lago.

c) Bill kicked Bob the ball. / Bill chutou a bola para o Bob.

Sobre a interação entre verbos e construções da estrutura argumental, é interessante observar que dados empíricos revelam alguns comportamentos de verbos que se diferem do que é previsto por teorias gramaticais categóricas. Por exemplo, o verbo do inglês *quiver* (tremar, sacudir) é considerado “intransitivo”, porém é encontrado em sentenças do tipo: “the Bird sat, quivering its wings” (ATKINS & LEVIN, 1995 *apud* MANNING, 2003, p. 298). De fato, as línguas se mostram mais flexíveis e variáveis do que muitas vezes se assume. Uma solução possível para a sentença mencionada com o verbo *quiver* seria incluir uma nova “subcategorização” para esse verbo. Porém, dessa forma perderíamos a informação importante de que o verbo *quiver* ocorre raramente na forma transitiva, sendo que sua forma intransitiva, apontada tradicionalmente como “a correta”, é a que representa o maior número de ocorrências desse verbo. Por outro lado, tais aspectos sobre o uso do verbo *quiver* podem ser capturados através da abordagem probabilística para estudos linguísticos.

Ilustremos como a abordagem probabilística lida com as possibilidades de interação entre verbos e padrões sintáticos através de outro exemplo dado por Manning (2003). Em uma investigação sobre o uso do verbo *regard*, o autor objetivou calcular a probabilidade de esse verbo ocorrer em diferentes padrões sintáticos. Para tal, ele analisou 300 ocorrências de *regard* retiradas do *The New York Times*, contou quantas vezes o verbo ocorreu em cada padrão sintático identificado na amostra analisada e dividiu o resultado pelo total de expressões de sua amostragem, ou seja, por 300. Desta forma, ele obteve a probabilidade da ocorrência de *regard* com os diferentes padrões frasais. Fica claro, portanto, que na perspectiva da Linguística Probabilística, “regras” gramaticais são substituídas pelas probabilidades de uso de um padrão ou

de associação entre esse padrão e fatores linguísticos e não-linguísticos. Captura-se, assim, o que é linguisticamente provável, ao invés do que é linguisticamente possível ou gramatical (CHATER & MANNING, 2006, p. 335).

Com base no exposto acima, observa-se que, assim como a Linguística de *Corpus*, a Linguística Probabilística considera o uso da língua em sua totalidade e possibilita descrições mais completas sobre fenômenos linguísticos. A partir dessa compreensão, seguimos Manning (2003) e adotamos uma perspectiva probabilística para nosso estudo sobre construções usadas para expressar transferência de posse no português, no inglês e na interlíngua português/inglês. Outra razão para recorrermos à Linguística Probabilística é o fato de essa ser apontada como uma abordagem ideal para investigar como falantes escolhem recursos de suas línguas para se comunicarem (Ibidem), como nos propomos a fazer nesta tese ao perguntarmos “O que influencia a escolha por uma das construções dativas?”. Vale a pena salientar que incluir informações acerca de preferências gramaticais, tradicionalmente vistas como parte da performance, em descrições linguísticas, significa expandir o escopo da gramática para nela incluir fatores semânticos e discursivos. Ademais, passa-se a admitir que grande parte dos exemplos de variabilidade na gramática fazem parte da competência, ao invés de serem simplesmente atribuídos à performance (Ibidem, p. 308).

Nossa opção por uma abordagem probabilística para nosso estudo fundamenta-se também no reconhecimento crescente de que a cognição humana baseia-se em processamentos probabilísticos (BOD, HAY & JANNEDY, 2003). Por essa razão, compreendemos que perspectivas cognitivas da língua devem investigar o papel de probabilidades na linguagem humana, vista nessa perspectiva como não-autônoma em relação a outras habilidades cognitivas. Pesquisas recentes nesse âmbito têm mostrado que as probabilidades - números que variam entre 0 (impossibilidade) e 1 (certeza) representando as chances de ocorrência de um dado evento em uma série de observações (BOD, 2003) - exercem de fato um papel central na aquisição, na representação, na compreensão e na produção linguística (BOD, HAY & JANNEDY, 2003).

Já foi constatado, por exemplo, que o processamento e a representação de palavras são fortemente influenciados pelas frequências de itens lexicais, que probabilidades de co-ocorrência de palavras são capturadas por falantes, que as probabilidades de sentenças ocorrerem podem ser previstas pelas probabilidades combinadas das subpartes das sentenças, e que alternâncias entre

construções linguísticas apresentam propriedades de contínuo e comportamento gradiente (Ibidem). As probabilidades se fazem presentes também em processos de desambiguação de expressões, tão comuns na experiência linguística dos falantes. Por exemplo, as probabilidades com que um verbo pode interagir com diferentes construções da estrutura argumental afetam a resolução de ambiguidade de sentenças (JURAFSKY, 2003). Em tarefas desse tipo, falantes tentam calcular, com base em distribuições probabilísticas, o mapeamento entre forma e significado condicionado pelo contexto em que a expressão ocorre. Resultados como os citados neste parágrafo mostram que os elementos básicos da teoria de probabilidades, os efeitos de frequência, atuam no sistema linguístico.

No que tange à aquisição, Pierrehumbert (2003) defende que adicionar probabilidades à Linguística torna o problema de aquisição mais fácil. Segundo o autor, com a atribuição de probabilidades às estruturas linguísticas, a aquisição dependeria apenas de evidência positiva do insumo. Sua argumentação é a de que padrões linguísticos seriam aprendidos com base em inferências estatísticas que se tornariam cada vez mais robustas com o aumento do tamanho da amostragem (insumo linguístico). Sub-representações estatísticas funcionariam como evidência negativa.

A proposta de um sistema linguístico probabilístico leva a reformulações de alguns conceitos da Linguística para aqueles que adotam a Linguística Probabilística. Em relação à noção de gramaticalidade, no lugar da distinção clara entre expressões julgadas categoricamente como gramaticais ou agramaticais, postula-se um contínuo de boa formação no qual algumas expressões são mais preferidas, algumas são menos frequentes e algumas não são usadas (BOD, HAY & JANNEDY, 2003, p. 5). A competência linguística, por sua vez, antes vista como consistindo em um conjunto mínimo de regras e de restrições categoriais, passa a incluir propriedades distribucionais, gradiência e variabilidade.

Para enfatizarmos o contraste entre abordagens categoriais e não categoriais, vejamos como cada uma delas trataria a seguinte ocorrência de *as least as* encontrada em um jornal: “*By the time their son was born, though, Honus Whiting was beginning to understand and privately share his wife’s opinion, as least as it pertained to Empire Falls*” (RUSSO, 2001 *apud* MANNING, 2003, p.292). Se, por um lado, a tradição linguística poderia considerar essa expressão como o resultado de um “erro”, por outro lado, abordagens probabilísticas

investigariam em outros textos se há mais ocorrências de “*as least as*”, como Manning (2003) fez. Após analisar textos dos anos 1994, 1995 e 2000, ele levantou a hipótese de que, apesar de muito menos comum que *at least as*, *as least as* indica uma mudança linguística em desenvolvimento, em direção ao uso dessa forma quando seguida por <as adjetivo as>, como na seguinte sentença: “*Steven P. Jobs has reemerged as a high-technology captain of industry, as least as far as the stock market is concerned*” (Ibidem, p. 293). Observa-se, portanto, que linguistas que assumem propriedades probabilísticas da língua buscam motivações para os dados empíricos que “fogem do esperado” antes de considerá-los “agramaticais” ou “falhas de produção”.

Mudanças proporcionadas por uma concepção probabilística da linguagem também têm ocorrido no que tange à noção de restrições linguísticas, as quais passam a serem compreendidas como generalizações estatisticamente robustas (PIERREHUMBERT, 2003). Manning (2003) explica que tais generalizações variam entre leves e rígidas de uma língua para a outra de acordo com as probabilidades a elas associadas em cada língua. Para ilustrar a gradiência das restrições linguísticas, o autor cria um evento hipotético no qual um policial o adverte. Tal evento poderia ser relatado por ele através de duas expressões semanticamente semelhantes: “O policial me advertiu” (construção ativa) e “Eu fui advertido pelo policial” (construção passiva). As seguintes restrições operantes no mapeamento entre o *input* semântico hipotético e sua expressão sintática são colocadas pelo autor (Ibidem, p. 318): (I) é preferível que o sujeito de uma sentença expresse o agente da ação descrita na mesma (restrição de mapeamento); (II) é preferível que a informação dada discursivamente (informação mais velha) seja o sujeito (restrição discursiva); (III) é preferível que participantes de primeira ou segunda pessoa sejam o sujeito (restrição de pessoa). Se a probabilidade da ativação da restrição (I) em uma língua é de 100%, a passiva nunca ocorrerá. Já no caso do inglês, nenhuma dessas restrições é categorial, de forma que ambas as construções podem ser utilizadas. Não obstante, tais restrições não deixam de atuar no inglês como restrições leves na seleção de uma das construções em questão, como mostram Bresnan, Dingare, and Manning (2001 *apud* Manning, 2003). Através da análise do *corpus* Switchboard⁴, os autores investigaram a influência da restrição de pessoa na seleção da construção passiva e constataram que a frequência da passiva com 1ª e 2ª pessoas agindo sobre terceira pessoas é significativamente menor, enquanto que a frequência da passiva com 3ª pessoas agindo sobre 1ª e

⁴ <http://catalog.ldc.upenn.edu/LDC97S62>

2ª pessoas é consideravelmente maior. Assim, ao invés de ditarem o que é e o que não é possível, as restrições passam a apontar o que é mais ou menos provável em uma língua.

Até agora, para falarmos sobre a Linguística de *Corpus* e a Linguística Probabilística, temos usado os termos “estruturas” e “padrões” da língua. Na próxima seção, explicitamos ainda mais a nossa compreensão sobre a natureza dessas “estruturas”, focalizando em especial aquelas que analisamos nesta tese.

2.3 Abordagem Construcionista

Uma vez que nos propomos a analisar aspectos da expressão linguística de eventos de transferência de posse, perguntamo-nos quais elementos de uma sentença nos indicariam que elas estão expressando esses tipos eventos. Intuitivamente, pode-se pensar que os verbos são os melhores indicadores. Porém, exemplos como os abaixo (GOLDBERG, 2003) mostram que tal intuição pode nos levar a um número limitado de dados:

7) Ocorrências do verbo *slice* (fatiar) em diferentes padrões sintáticos:

- a) *He sliced the bread.*
- b) *Pat sliced the carrots into the salad.*
- c) *Pat sliced Chris a piece of pie.*

O verbo “slice” evoca, prototipicamente, um frame semântico (GOLDBERG, 2010) que não é o de transferência de posse. De fato, a interpretação geral do evento em (7a) é a de uma entidade agindo sobre a outra, e em (7b) a de uma entidade fazendo com que outra entidade se mova até um local. Não obstante, (7c) expressa um evento de transferência de posse, um evento de transferência de um pedaço de torta de Pat para Chris, se quisermos ser mais específicos. Se o mesmo verbo, *slice*, é utilizado nas três sentenças acima, mas em cada uma delas visualizamos tipos de eventos diferentes, chegamos à conclusão de que o verbo utilizado não é o suficiente

para interpretarmos o evento geral sendo descrito por uma sentença. Que outro fator, então, seria também responsável pelas diferentes interpretações dos eventos das sentenças acima? Encontramos uma resposta na Gramática de Construções, segundo a qual padrões sintáticos básicos como os acima são diretamente associados a estruturas semânticas (Ibidem), como ilustra o Quadro 3.

Quadro 3 - Associações entre padrões sintáticos e estruturas semânticas

Forma	Interpretação geral
<Suj. Verb. Obj.>	X age sobre Y
<Suj. Verb. Obj. Obl.>	X faz com que Y se mova até Z
<Suj. Verb. Obj.1 Obj.2>	X faz com que Y receba Z

Assim, entendemos que os sentidos das expressões 7a-c acima são construídos a partir da interação entre o verbo *slice* e os padrões sintáticos nos quais o verbo ocorre, os quais contribuem com informações semânticas sobre a interpretação geral do evento sendo descrito.

Pareamentos entre formas e funções semânticas como os do Quadro 3 são chamados de construções (ELLIS & FERREIRA-JUNIOR, 2009; GOLDBERG, 1995). Outros exemplos de construções fornecidos no Quadro 4 mostram que, para a Gramática de Construções, construções fazem parte de todos os níveis de análise gramatical, desde morfemas até padrões sintáticos, e que, em alguns casos, além de funções semânticas, elas possuem também funções discursivas e pragmáticas.

Quadro 4 - Exemplos de construções de diferentes níveis de representação linguística

Construção	Forma	Funções
Morfema	Anti-	Sufixo que significa ação contrária, oposição, contrariedade.
Palavra	Careca	Palavra informal que se refere a um indivíduo que não possui cabelo na cabeça ou em parte dela e que pode ter tom pejorativo.
Expressão idiomática	Mamão com açúcar.	Expressão que se refere a uma tarefa que pode ser realizada facilmente.
Construção Passiva	Subj aux VPpp (PPby) A janela foi quebrada pelo menino.	Função discursiva: topicalizar paciente.

Adaptado de Goldberg (2003)

Observamos na descrição da função da construção “careca” acima, por exemplo, que fatos sobre o uso das construções (registro, variedade linguística etc.) são levados em consideração na caracterização das mesmas, em consonância com o que defende a Linguística de *Corpus*.

À luz da perspectiva construcionista, podemos dizer que o nosso interesse de pesquisa são construções associadas à estrutura semântica de transferência de posse. No seu pólo formal, tais construções devem especificar três posições sintáticas, correspondentes ao número de participantes semânticos (quem transfere - agente, o que transfere - tema, para quem é transferido - recipiente) envolvidos em eventos de transferência de posse (GOLDBERG, 2006, p. 189). Em inglês, essas construções são:

Quadro 5 - Construções da estrutura argumental do inglês utilizadas na expressão de eventos de transferência de posse

Construção	Pólo formal	Exemplo
Ditransitiva	<Sujeito__agente Verbo Objeto1__recipiente Objeto2__tema>	<i>Please give this matter your attention.</i> (Extraído do subcorpus <i>Written</i>)
Ditransitiva Preposicionada	<Sujeito__agente Verbo Objeto__tema Oblíquo__recipiente >	<i>I've just finished showing the tape to Julie.</i> (Extraído do subcorpus <i>Written</i>)

Devemos ressaltar que a “transferência de posse” que nos interessa inclui não apenas expressões que denotam a “transferência literal de um objeto físico de um agente para um recipiente”, como nos exemplos do Quadro 5. Ao falarmos sobre expressões de transferência de posse ao longo desta tese, referimo-nos também a sentenças que expressam algum outro aspecto saliente da estrutura semântica de eventos de transferência ou que envolvem metáforas. Discutimos cada um desses casos abaixo.

No uso da língua, construções da estrutura argumental interagem com construções lexicais e com outras construções gramaticais para formarem sentenças. Essa possibilidade de combinação entre construções explica, segundo Goldberg (2006, p. 22), a criatividade linguística. Porém, a mesma autora aponta que a combinação entre construções não ocorre de forma aleatória. Antes, para que construções linguísticas possam ser combinadas, suas propriedades precisam ser semanticamente compatíveis. No caso da interação entre verbos e construções da estrutura argumental, é necessário que os primeiros designem algum aspecto saliente da estrutura semântica das últimas (Ibidem, 1995, p. 65). Por exemplo, verbos que ocorrem na construção ditransitiva, diretamente associada à semântica de transferência de posse, podem designar uma pré-condição para a transferência (*I baked you a cake*) ou o modo utilizado para tornar a mesma possível (*She threw him the ball*). Além disso, os eventos designados pelo verbo e pela construção da estrutura argumental devem compartilhar pelo menos um papel semântico em comum. Se, por um lado, a estrutura semântica de um verbo especifica “papéis participantes” particularmente associados a ele, por outro lado, a semântica associada a uma construção da estrutura argumental delimita “papéis argumentais” de natureza mais geral. A compatibilidade semântica entre um verbo e uma construção sintática é alcançada quando os papéis participantes do primeiro podem ser interpretados como exemplos dos papéis argumentais da segunda. A interação entre o verbo *hand* (passar) e a construção ditransitiva ilustra total correspondência entre papéis participantes e papéis argumentais: os papéis participantes <*hander*>, <*handee*> e <*handed*> são exemplos dos papéis argumentais <agente>, <recipiente> e <tema>, respectivamente. Em casos menos prototípicos, nos quais o verbo possui apenas dois papéis participantes, por exemplo, a mesma coerência semântica deve ser mantida. Como ilustração, o verbo *kick* (chutar) possui dois papéis participantes, <*kicker*> e <*kicked*>, que são exemplos dos papéis argumentais <agente> e <tema> da construção ditransitiva. Em sentenças como “I kicked him the ball”, a construção ditransitiva contribui com o papel argumental <recipiente> para a

interpretação do evento. Por outro lado, se um verbo possui três papéis participantes, mas um deles entra em conflito semântico com os papéis argumentais de determinada construção sintática, a sentença resultante da interação entre estas duas construções tende a não ser aceita. Por exemplo, em **I put the table flowers*, há conflito entre o papel argumental <recipiente> da construção ditransitiva e o papel participante <lugar colocado> do verbo *put* (colocar).

Para haver compatibilidade entre construções lexicais e a construção ditransitiva, Goldberg (1995) lista alguns “princípios” que devem ser respeitados. À luz da Linguística Probabilística, apresentada na seção anterior, interpretamos tais “princípios” como generalizações feitas a partir da observação da frequência de interação entre a construção ditransitiva e construções lexicais, que podem estar mais ou menos ativa em diferentes variedades no inglês:

a) A construção ditransitiva não interage com verbos incompatíveis com a noção de transferência. Devido ao princípio de coerência semântica, os papéis participantes e os papéis argumentais não podem ser conflituosos. **She removed Sally the garbage*.

b) O argumento <agente> tem a intenção de transferir algo através de sua ação, ou seja, precisa agir com “controle” (CANÇADO, 2005). A sentença “*Bob told Joe a story*” não pode significar, por exemplo, que Joe ouviu a história acidentalmente enquanto Bob a contava para outra pessoa.

c) O argumento <recipiente> é tipicamente animado, o que explica a não-naturalidade de **Mary brought the table flowers* (<table> não possui o traço [animado]). Além disso, ele é entendido como um beneficiário ou alguém disposto a receber algo: **Sally burned Joe some rice although he had said he hates it*.

d) A construção ditransitiva implica que o argumento designado pelo primeiro objeto possa receber o argumento designado pelo segundo objeto: **Joe cleared Sam the floor*.

Expressões com a construção ditransitiva nas quais as generalizações supracitadas aparentemente não estão operantes envolvem metáforas. Entendemos “metáfora” como um processo cognitivo através do qual um modelo ou esquema de um domínio conceitual (domínio fonte) é projetado para um modelo ou esquema de outro domínio (domínio alvo). (PALMER,

1996). Por exemplo, na sentença “*Mary gave me the flu*”, “causar um efeito em uma entidade” corresponde a “transferir esse efeito”, entendido metaforicamente como um objeto. O fator “controle”, porém, não é mapeado do domínio fonte para o domínio alvo nesse exemplo. Outras metáforas sistemáticas que possuem como domínio fonte a “transferência bem sucedida de um agente com iniciativa/controle para um recipiente pré-disposto” incluem: a) comunicação “viajando” do estímulo para o ouvinte: *She told Jo a fairy tale*; b) “percepções” como entidades que se movem em direção ao experienciador: *He showed Bob the view*; c) “ação intencionalmente dirigida a uma pessoa” como “entidade transferidas para a mesma”: *She gave him a punch*; d) “ação em benefício da pessoa” como “objeto transferido para a mesma”: *Cry me a river*.

Nota-se, a partir do exposto acima, que excluir expressões que envolvem metáforas de nossa análise significaria excluir muitos dados linguísticos. Afinal, as metáforas permeiam a linguagem humana (LAKOFF & JOHNSON, 1980). Através do conceito de metáfora conceitual, entendemos também uma distinção apontada por Goldberg (1995) entre a construção ditransitiva e a construção ditransitiva preposicionada. A autora explica que, enquanto a construção ditransitiva é diretamente associada ao sentido de transferência de posse independentemente do verbo com o qual interage, a ditransitiva preposicionada, em exemplos como *I propose to send your description to them* (Written), é uma extensão metafórica da construção de movimento causado (ver Quadro 1), licenciada pela metáfora “transferência de posse como transferência física”.

A diferença semântica supracitada entre as construções ditransitiva e ditransitiva preposicionada e a visível diferença sintática entre essas construções não são os únicos aspectos que as diferenciam. Compreender tais aspectos é fundamental para estudos que, assim como o nosso, procuram entender as motivações para o uso de uma delas quando a outra também seria possível. Exploramos essa questão na seção 2.5.1 desta tese, quando revisamos estudos anteriores sobre as construções dativas do inglês baseados em análises de corpora.

Nesta seção expusemos nossa perspectiva construcionista para a análise linguística que iremos fazer. Na próxima seção, destacamos como estudos sobre construções gramaticais em particular são beneficiados pela Linguística de *Corpus*.

2.3.1 Estudos sobre alternâncias entre construções baseados em análise de *corpus*

Segundo o “princípio da não sinonímia” (GOLDBERG, 1995 p. 67; GRIES, 2009b, p. 5), duas construções sintaticamente distintas devem ser distintas quanto a algum aspecto semântico, discursivo ou de uso. De fato, estudos baseados em corpora têm mostrado cada vez mais que construções não variam aleatoriamente nas produções dos falantes, antes são escolhidas sob a influência de fatores linguísticos (associações léxico-gramaticais, aspectos discursivos etc.) e não linguísticos (variedade linguística, registro etc.) (BIBER et al., 1998; KENNEDY, 1998). Tais estudos possibilitam descrições mais detalhadas que apontam diferenças às vezes sutis entre construções gramaticais que não poderiam ser observadas de outra forma.

Gries (2001), por exemplo, investigou a alternância entre as duas construções transitivas de “verbo-partícula” do inglês, ilustradas pelo par de sentenças abaixo:

8) Alternância entre construções transitivas de “verbo-partícula” do inglês:

(a) John picked up the book. – construção 0

(b) John picked the book up. – construção 1

O objetivo do autor foi responder as seguintes perguntas: Como cada variante se difere uma da outra? Por que e como diferentes fatores influenciam a escolha subconsciente por uma das construções transitivas de “verbo-partícula” em contextos reais de uso? Como as escolhas de falantes em situações discursivas específicas podem ser previstas? Sua hipótese inicial é a de que, uma vez que falantes esforçam-se para se comunicar com o menor custo de processamento possível, e a construção 0 é mais fácil de se processar (HAWKINS, 1994; ROHDENBURG, 1996 *apud* GRIES, 2001, p. 39), essa construção será selecionada em situações nas quais o custo de processamento associado à expressão já for alto. Por outro lado, se o sintagma verbal não requer muito esforço de processamento, a construção 1 tende a ser escolhida. Ao analisar o custo de processamento de expressões, Gries considera a contribuição de diversos fatores. Ele explica, por

exemplo, que quanto maior e mais complexo o objeto direto for, mais esforço e memória de trabalho são necessários para processar a expressão (Ibidem, p. 38), e que o custo de processamento será menor para objetos diretos que representam informações já dadas no discurso (GIVÓN, 1992 *apud* GRIES, 2001, p. 39). Outras variáveis analisadas foram a animacidade do objeto direto e a idiomaticidade do sintagma verbal.

Para testar a sua hipótese, Gries analisou construções de “verbo-partícula” presentes no British National Corpus (BNC). O autor compilou 403 sentenças do *corpus* - 200 da seção *Spoken* e 203 da seção *Written* – que continham a construção de verbo-partícula composta por verbos altamente frequentes. Após a análise dos fatores supracitados, o uso de testes estatísticos criteriosos tornou o estudo bastante confiável. Gries interpreta seus resultados como confirmadores de sua hipótese de processamento, já que as variáveis que indicam um baixo grau de custo de processamento, como sintagmas verbais literais com objetos diretos pronominais, por exemplo, favoreceram a ocorrência da construção¹, enquanto que as variáveis que indicam maior grau de custo de processamento, como referentes de sintagmas nominais novos discursivamente, favoreceram a ocorrência da construção 0. A análise feita apontou também a seguinte classificação de força de grupos de variáveis na escolha das construções: variáveis discurso-funcionais (contexto precedente), variáveis sintáticas, variáveis semânticas e outras variáveis. Com esse exemplo das construções de “verbo-partícula” do inglês, vimos como a Linguística de *Corpus* permite diferenciar construções gramaticais tão semelhantes.

Outro trabalho que busca identificar diferenças entre pares alternantes de construções gramaticais é o de Gries e Stefanowitsch (2008), os quais advogam a necessidade de se considerar variáveis como canal (escrito/falado), registro (formal/informal), variedade linguística e gênero dos falantes em pesquisas no domínio léxico-gramatical. Os pares de construções alternantes investigados pelos autores na análise do componente britânico do *International Corpus of English* (ICE-GB) foram (I) as construções ativa e passiva, (II) as construções de futuro com *will* e com *going to* e (III) as construções de verbo-partícula. Os resultados obtidos indicaram, entre outros, que enquanto algumas construções mostram-se sensíveis às diferenças de formalidade associadas aos canais “fala” e “escrita”, como as construções de futuro, outras construções não apresentaram essa característica. A construção passiva, por exemplo, é mais fortemente associada ao vocabulário formal e retém essa associação até mesmo em registros mais

informais típicos da fala.

Os dois estudos supracitados mostram como abordagens cognitivamente realistas baseadas no uso para o estudo da linguagem ampliam o conhecimento já existente sobre construções linguísticas. Através de estudos como esses, construções aparentemente sinônimas são cada vez mais diferenciadas em termos de suas associações com aspectos sintáticos, semânticos e discursivos, com diferentes registros, variedades linguísticas e gêneros específicos de certos grupos de falantes.

Para finalizarmos nossa revisão sobre os principais pontos de abordagens construcionistas que iluminam nossa perspectiva teórica, apresentamos a seguir uma proposta construcionista para a aquisição da língua.

2.3.2 Aprendendo construções

Concordamos com o argumento de que, se construções semi-idiossincráticas existem em todas as línguas e devem ser aprendidas com base no insumo linguístico, é plausível que os casos mais gerais, mais regulares e frequentes sejam também aprendidos dessa maneira por uma espécie que possui habilidades sociais complexas e capacidades cognitivas superiores⁵ (GOLDBERG, 2006). Por essa razão, adotamos uma perspectiva de aquisição da linguagem baseada no uso, segundo a qual as construções de uma língua são aprendidas a partir do insumo linguístico e através de processos cognitivos gerais (CROFT & CRUSE, 2004; ELLIS, 2003; ELLIS & LARSEN-FREEMAN, 2009; GOLDBERG, 1995; LANGACKER, 1987; PALMER, 1996; TOMASELLO, 2006).

⁵ Segundo Tomasello (2003b), essas capacidades cognitivas superiores incluem a compreensão de intencionalidade e de causalidade, isto é, a compreensão das forças mediadoras de eventos externos que não são observáveis de maneira direta. Outra característica apontada pelo autor como única da espécie humana é a capacidade de cada indivíduo compreender os co-específicos como seres com vidas mentais e intencionais iguais às dele, o que permite aprendizagens culturais por instrução e por colaboração.

Apesar de reconhecermos que a aquisição de uma língua é um fenômeno tão complexo quanto a espécie humana, envolvendo, dentre outros, fatores sociais e culturais (ELLIS & LARSEN-FREEMAN, 2009), focalizamos aqui sua dimensão cognitiva. Dentre os processos cognitivos envolvidos na aquisição da língua, podemos citar a lembrança de sentenças ouvidas, a formação de padrões a partir de estímulos recebidos, a analogia, a generalização de esquemas conceituais e a formação de protótipos a partir de exemplares (Ibidem). Vejamos abaixo como alguns desses processos cognitivos podem explicar a aquisição de construções da estrutura argumental em particular.

Existem evidências de que falantes retêm informações sobre o uso específico de itens lexicais (ELLIS, 1999; GOLDBERG, 2006; LANGAKER, 1987; TOMASELLO, 2000). Ao testemunhar expressões como *I've told you the truth (Written)*, por exemplo, um aprendiz de inglês pode guardar informações sobre o ambiente sintático no qual o verbo *tell* é utilizado. Memorizar sentenças ouvidas, contudo, não é o suficiente para interagirmos em novas situações comunicativas, nas quais precisamos prever significados com base nos itens lexicais e nas estruturas gramaticais testemunhadas (compreensão) e escolher itens lexicais e estruturas gramaticais de acordo com a mensagem que queremos transmitir (produção). Necessidades como essas levam à formação de categorias linguísticas (GOLDBERG et al., 2007).

A categorização, que serve como base para muitos processos mentais importantes, tais como a percepção, a lógica e a aprendizagem em geral, é um processo cognitivo, automático e inconsciente que consiste no agrupamento de exemplares que compartilham alguma semelhança observável (LAKOFF, 1987). Assim, após aprenderem de forma conservadora expressões linguísticas concretas testemunhadas no insumo linguístico, falantes usam suas habilidades cognitivas para categorizar e esquematizar tais expressões e estruturas, abstraindo categorias que lhes possibilitem agir linguisticamente (falar ou escrever) (TOMASELLO, 2000). No caso das construções da estrutura argumental, a formação de categorias é motiva porque, como já vimos, essas construções são muito úteis para prevermos o sentido geral de uma expressão.

No aprendizado do pólo formal das construções da estrutura argumental, padrões sintáticos emergem a partir de generalizações feitas a respeito dos ambientes sintáticos nos quais os verbos ocorrem. A coleção de memórias de sentenças testemunhadas permite com que exemplares semelhantes (ex: *I'll give you the school's number, He has recently sent me his advice*

etc.) sejam relacionados através de processos de analogia que resultam na emergência de categorias linguísticas abstratas (ex: <Suj. Verb. Obj.1 Obj.2>) (ELLIS, 2002). Já o aprendizado do pólo semântico da construção da estrutura argumental se dá a partir da associação entre a estrutura sintática (pólo formal da construção) e os significados dos verbos que nela ocorrem mais frequentemente. Evidências disso são fornecidas por Goldberg, Casenhiser e Sethuraman (2003) em seu estudo sobre a aquisição de inglês/L1. Os pesquisadores, ao analisarem transcrições de conversas entre 27 crianças com 28 meses de idades e suas respectivas mães em três tipos de interação (brincadeiras, contação de histórias e hora do lanche), constataram que existe uma forte tendência de um verbo específico ser usado pelas mães com grande frequência em cada construção da estrutura argumental analisada. Isso permite, segundo os autores, que as crianças fixem significados para as estruturas sintáticas com base nos significados dos verbos nelas utilizados (Quadro 6). Semelhantemente, a mesma tendência de um único verbo ser usado com mais frequência em cada construção da estrutura argumental foi encontrada nas falas das crianças, o que mostra como elas são sensíveis ao insumo linguístico que recebem.

Quadro 6- Significados compartilhados entre construções da estrutura argumental e os verbos mais frequentes em cada uma delas

Verbo	Significado compartilhado entre o verbo e a construção	Construção
Put	X causes Y to move Z	Construção de movimento causado
Go	X moves Y	Construção de movimento intransitive
Give	X causes Y to receive Z	Construção Ditransitiva

Adaptado de Goldberg et al. (2003)

Goldberg e colegas (Ibidem, p. 15) explicam que a associação entre forma e significado no aprendizado de construções da estrutura argumental pode ocorrer também a partir de verbos distintos usados menos frequentemente na mesma construção, requerendo apenas uma abstração mínima da semântica comum aos verbos testemunhados no mesmo padrão frasal.

A proposta de aquisição apresentada acima se aplica também à aquisição de construções da estrutura argumental de uma segunda língua, como afirma Ellis (2002, p. 170), e é adotada nesta tese ao tratarmos da aquisição de inglês por aprendizes brasileiros. Evidências da aquisição

de construções de uma segunda língua serão citadas mais à frente. Antes, porém, as especificidades da natureza e do desenvolvimento do conhecimento sobre a L2 são discutidas na próxima seção.

2.4 Interlíngua

Chamamos de Interlíngua o sistema linguístico desenvolvido por um aprendiz durante a aquisição da L2 (CORDER, 1993; ELLIS, 1997; SELINKER, 1972). A partir de uma perspectiva construcionista, compreendemos a Interlíngua como reflexo da construção do conhecimento da L2 por parte de um aprendiz que, através de suas habilidades cognitivas gerais, desenvolve um sistema dinâmico que sofre mudanças na medida em que a experiência com a L2 aumenta.

O conceito original de Interlíngua (SELINKER, 1972) trouxe duas importantes contribuições para a pesquisa em aquisição de segunda língua. Uma delas foi o reconhecimento da língua do aprendiz como um sistema próprio, independente da L1 e da L2. A outra diz respeito ao reconhecimento da coerência e da lógica existentes nesse sistema, de forma que a noção de “erro” (produção que se desvia das normas da língua-alvo) foi revista. Produções da interlíngua antes consideradas “desvios” em relação à língua-alvo passaram a ser vistas como evidências de um sistema internamente estruturado, que pode possuir elementos próprios, não apenas aqueles provenientes da língua materna do aprendiz e da língua-alvo (GASS & SELINKER, 1994).

Para explicar a existência de elementos que se distanciam das regras da língua-alvo e que podem permanecer na Interlíngua ao longo do tempo, Selinker (1972) propõe cinco processos psicolinguísticos: (I) supergeneralização de regras da língua alvo (como no uso do sufixo -ed em “He goed to the movies”); (II) transferência de treinamento (resultado da influência de procedimentos de ensino); (III) estratégias de aprendizado para lidar com o conteúdo a ser aprendido (como a simplificação de regras da língua-alvo); (IV) estratégias de comunicação (como cessar o aprendizado de novas formas por considerar que o conhecimento já adquirido é suficiente para a comunicação com falantes nativos da língua-alvo); (V) transferência linguística (uso de elementos da L1 na performance em L2).

Embora tenha contribuído para os estudos em aquisição de segunda língua, a noção

original de Interlíngua passou a ser criticada a partir do final da década de 90, quando linguistas como Firth e Wagner iniciaram uma oposição à consideração exclusiva de processos psicolinguísticos em estudos sobre a Interlíngua (KRAMSCH & WHITESIDE, 2007). Desde então, foco também tem sido dado à dimensão social da aquisição de uma segunda língua, o que levou a novas interpretações sobre fenômenos interlinguais. Por exemplo, a mudança de código (code-switching) da L2 para a L1 durante a comunicação na língua-alvo, antes interpretada como um recurso utilizado pelo aprendiz para evitar formas da L2 ainda não conhecidas por ele, passou a ser vista como uma possível estratégia para marcar a identidade do aprendiz. A transferência linguística, por sua vez, passou a incluir não apenas a transferência de formas linguísticas, mas também a transferência de mapeamentos conceituais (JARVIS, 2011), de estilos conversacionais (*conversational styles*) e de generalizações de uso pragmático (BARRON, 2003).

De fato, os processos de aquisição e de uso de uma segunda língua são fenômenos bastante complexos, influenciados por fatores de diversas naturezas. Devido ao escopo deste trabalho, discutiremos aqui apenas a influência da língua materna (CORDER, 1993; ELLIS, 1985; KELLERMAN, 1979; LAKSHMANAN & SELINKER, 2001; ODLIN, 1989; SCHACHTER, 1993; SELINKER, 1992;). Segundo Odlin (1989), essa influência pode fazer com que aprendizes: (I) evitem o uso de estruturas da L2 que sejam diferentes das estruturas de sua L1; (II) usem alguns elementos da L2 com maior frequência do que falantes nativos da língua-alvo; (III) compreendam mensagens na L2 de forma equivocada devido a diferenças fonéticas, sintáticas ou culturais entre a L1 e a L2; (IV); usem formas da L1 na performance da L2. Quando o uso da língua-alvo se dá de uma maneira que é semântica ou sintaticamente apropriada para a L1, mas não para a língua-alvo, diz-se que ocorre uma transferência negativa. Já o uso de formas da L1 que resultam em “acertos” no desempenho na L2 é chamado de transferência positiva.

Embora a ocorrência de transferência linguística seja amplamente reconhecida e documentada, existem diferentes propostas acerca de sua natureza (LAKSHMANAN & SELINKER, 2001). Por exemplo, Corder (1978) defende que a transferência é uma estratégia comunicativa, ocorrendo quando aprendizes recorrem à sua L1 para suprirem a sua falta de conhecimento sobre a língua-alvo em situações comunicativas. Para o autor (1993, p. 25), a transferência consiste em um processo que se dá da estrutura mental que é o conhecimento implícito sobre a L1 para o sistema independente da interlíngua. Por outro lado, Schachter (1993,

p. 32) afirma que “transferência” corresponde a testes de hipóteses feitos pelos aprendizes a respeito do funcionamento da L2 com base em sua L1.

Seja qual for a natureza da transferência de elementos da L1 para a Interlíngua, esse fenômeno não parece ocorrer livremente. Antes, existem condições que inibem ou favorecem a transferência linguística. Dentre elas, destacamos:

- (I) O nível de proficiência do aprendiz: quanto menor o nível de proficiência do aprendiz, maiores são as chances de ocorrer transferência negativa (Odlin, 1989). Com o aumento do nível de proficiência, contudo, a transferência conceitual pode aumentar, já que o aprendiz possui mais recursos da L2 para expressar conceitos de sua L1 (JARVIS, 2000).
- (II) A quantidade de exposição e uso da L2: a transferência linguística tende a diminuir com o aumento da exposição e do uso da língua-alvo (MURPHY, 2003).
- (III) A psicotipologia do aprendiz: a transferência linguística é favorecida quanto maior a percepção da proximidade tipológica entre a L1 e a L2 por parte do aprendiz (KELLERMAN, 1983). Em outras palavras, quanto maior a proximidade entre a L1 e a L2 do ponto de vista do aprendiz, maiores as chances de ocorrer transferência.
- (IV) Efeitos de frequência: construções frequentes na L1 apresentam maior probabilidade de serem transferidas de forma não-intencional durante produções da Interlíngua (POULISSE & BONGAERTS, 1994).

A partir de uma perspectiva probabilística da língua, sugerimos que a transferência da L1 para a interlíngua é resultado da ativação de padrões probabilísticos da L1 influenciada pela interação dos fatores acima.

Agora que já definimos o que entendemos por interlíngua, conciliamos na próxima seção de forma mais explícita essa hipótese com as abordagens construcionistas previamente discutidas, apontando evidências da existência de construções na Interlíngua.

2.4.1 Evidências da existência de construções na Interlíngua

Gries e Wulff (2005) relatam resultados que apontam para a presença de construções sintáticas na interlíngua alcançados por meio de diferentes metodologias. Por exemplo, através de uma tarefa de completar sentenças com as construções dativas do inglês, realizada por falantes nativos do alemão, os autores encontraram efeitos de *priming* sintático (tendência para a repetição de estruturas sintáticas usadas anteriormente). O fato de aprendizes exibirem efeitos de *priming*, assim como falantes nativos da língua-alvo, é interpretado como evidência de que eles possuem representações dos polos formais das construções da estrutura argumental (Ibidem, p. 185). Para investigarem a presença de representações dos polos semânticos das construções, Gries e Wulff replicam o experimento de Bencini e Goldberg (2000 *apud* GRIES & WULFF, 2005, p. 191) de classificação de frases, as quais poderiam ser agrupadas segundo semelhanças semânticas entre os verbos ou entre os padrões sintáticos. Os resultados obtidos revelam que aprendizes reconhecem sentenças como mais semelhantes quando elas exibem as mesmas estruturas frasais, o que é interpretado como evidência de que aprendizes possuem representações sobre as semânticas das construções.

Outro estudo que aponta para a validade de construções na interlíngua é o de Zara (2009). A autora utilizou uma tarefa de julgamento de aceitabilidade de sentenças com o objetivo de identificar em qual nível de proficiência em inglês/L2 (iniciante, intermediário, avançado) aprendizes brasileiros adquirem a construção ditransitiva do inglês. Nessa tarefa, ela incluiu expressões nas quais a construção ditransitiva foi apresentada com verbos inventados, cujos significados propostos envolviam a transferência de posse, para verificar se aprendizes seriam capazes de estender o uso da construção ditransitiva para novos itens lexicais, demonstrando que já fizeram generalizações acerca da estrutura argumental (WHITE, 2003, p. 210). Os resultados de Zara mostraram que um número expressivo de aprendizes do nível avançado considerou aceitáveis as expressões com os verbos inventados, o que indica que a construção ditransitiva do inglês já existe como uma categoria abstrata nas interlínguas desses aprendizes.

Resultados de estudos como os supracitados nos permitem adotar uma abordagem construcionista para o estudo da Interlíngua. Apesar de existirem poucos dados longitudinais coletados que registram o desenvolvimento do conhecimento sobre construções da L2, as

evidências já disponíveis corroboram a seguinte sequência desenvolvimental na aquisição de construções sintáticas da L2, que se assemelha à proposta para a aquisição da L1: (I) o aprendizado se inicia com a memorização de sequências de palavras (fórmulas); (II) quando vários exemplos de fórmulas já estão disponíveis na memória de longo prazo, estas são analisadas através de processos implícitos com base nos quais suas características comuns são identificadas; (III) esquemas mais abstratos são formados a partir de protótipos que emergem como tendências centrais nas categorias linguísticas (ELLIS, 2003).

Discorreremos até agora sobre as bases teórico-metodológicas que fundamentam esta tese. Nas próximas seções deste capítulo, revisaremos os estudos anteriores que nos informam sobre as construções dativas no inglês, na interlíngua e no português. Foram esses estudos que nos instigaram a levantar os questionamentos desta pesquisa.

2.5 Estudos anteriores sobre construções dativas

Eventos de transferência física ou metafórica de posse fazem parte das cenas mais básicas da experiência humana (GOLDBERG, 1995). Uma vez que as possibilidades de expressão linguística desses eventos apresentam variações translinguísticas e intralinguísticas notáveis, linguistas de diferentes correntes teóricas têm se interessado em investigar as chamadas “construções dativas” em diversas línguas. Vejamos nesta seção o que alguns estudos já realizados nos revelam sobre diferentes aspectos relacionados às construções dativas presentes em algumas variedades do inglês, do português do Brasil e de interlínguas de aprendizes de inglês de diferentes línguas maternas.

Durante nossa leitura sobre estudos anteriores acerca das construções dativas, buscamos resultados de estudos baseados em corpora pelo fato de eles apresentam as vantagens apontadas na seção 2.1 desta tese. Revisamos também alguns estudos que seguiram outras orientações teórico-metodológicas, seja porque seus resultados também são relevantes (no caso do inglês), ou porque estudos baseados em corpora representativos foram encontrados em pequenos números

(no caso da interlíngua) ou não foram encontrados (no caso do português).

2.5.1 Estudos anteriores sobre construções dativas no inglês

Em inglês, existem 3 construções tipicamente utilizadas para se falar sobre a transferência de posse: to-dative (*I gave a book to Mary*), for-dative (*I made a cake for Mary*), chamadas de formas dativas/ditransitivas preposicionadas, e a dativa de objeto duplo, ou ditransitiva (*I gave Mary a book*) (TOMASELLO, 2003a, p. 121). Alguns estudos baseados na análise de corpora que investigaram os fatores linguísticos e extra-linguísticos que podem influenciar a escolha por uma dessas construções são revisados a seguir.

Pautando-se na teoria de protótipos, Gries (2003) descreveu os casos prototípicos de cada construção que faz parte da alternância dativa do inglês (AD) através da análise de dados do *British National Corpus* (BNC). Para isso, o autor identificou o grau de relação entre as construções dativas e as variáveis listadas no Quadro 7, já apontadas na literatura como fatores que influenciam a seleção de uma dessas construções.

Quadro 7 - Variáveis que influenciam a seleção de construções dativas apontadas na literatura

Variável	Valor para a construção ditransitiva	Valor para a construção ditransitiva preposicionada
Processo descrito	NP _{sujeito} transfere intencionalmente NP _{tema} para NP _{recipiente}	
Animacidade do referente do NP _{recipiente}	Animado / humano	Não animado / menos animado
Animacidade do referente do NP _{tema}	Inanimado / não-humano	
Extensão ⁶ do NP _{recipiente}	Curto	Longo
Extensão do NP _{tema}	Longo	Curto
Determinante do NP _{recipiente}	Definido	Indefinido
Tipo de NP do NP _{recipiente}	Pronominal	Lexical
Determinante do NP _{tema}	Indefinido	Definido
Tipo de NP do NP _{tema}	Lexical	Pronominal
Novidade do referente do NP _{recipiente}	Dado (given)	Novo
Novidade do referente do NP _{tema}	Novo	

Adaptado de Gries (2003, p. 9)

Após selecionar os 10 verbos mais frequentes que participam da AD e extrair 117 construções dativas do *corpus* (60 construções ditransitivas e 57 construções ditransitivas preposicionadas), Gries utilizou técnicas estatísticas para verificar a relação entre cada variável acima (variáveis independentes) e cada uma das construções dativas (variáveis dependentes). Os resultados obtidos levaram às seguintes conclusões sobre os exemplares prototípicos de cada construção:

⁶ Gries (2003) explica que “curto” e “longo” referem-se ao número de palavras e sílabas dos argumentos. No exemplo “The Iran/Contra scandal gave the French much cause for amusement”, ele define o recipiente como “curto” por ser composto por 2 palavras monossílabas e o tema como longo por ser composto por 4 palavras que totalizam 6 sílabas.

a) Construção ditransitiva preposicionada prototípica: a sentença não denota transferência intencionada; o argumento <tema> é curto, o argumento <recipiente> é longo, complexo e não prototípico (não consciente, não voluntário); o argumento <tema> é indefinido, o argumento <recipiente> é definido.

b) Construção ditransitiva prototípica: a sentença denota transferência intencionada de um NP (*noun phrase* / sintagma nominal) <tema> para um NP<recipiente>, o argumento <recipiente> é curto, pronominal, animado e dado no discurso; o argumento <tema> é comparativamente mais longo e complexo.

Dentre as variáveis indicadas no Quadro 7, Gries destaca as propriedades do argumento <recipiente> e as variáveis morfossintáticas em geral como as que possuem maior poder discriminatório entre as duas construções dativas.

Em trabalho posterior, Gries e Stefanowitsch (2004) adotam uma abordagem semântica, segundo a qual a interação entre um verbo e uma construção é licenciada ou não pela compatibilidade entre a semântica do verbo e a semântica da construção (GOLDBERG, 1995), e investigam se a escolha por uma das construções que fazem parte da AD pode ser prevista apenas com base no verbo utilizado. Através da análise do componente britânico do *International Corpus of English* (ICE-GB), os autores identificaram colexemas distintos, ou seja, lexemas que exibem preferência por uma determinada construção, interagindo frequentemente com a mesma. Com base em resultados obtidos através de procedimentos estatísticos, os autores concluíram que a construção ditransitiva ocorre com maior frequência com verbos que denotam alguma forma de transferência (literal ou metafórica) que envolve contato direto entre o <agente> e o <recipiente>, enquanto que a construção ditransitiva preposicionada é mais comum com verbos que envolvem alguma distância entre o <agente> e o <recipiente>, como se o <tema> precisasse se mover por um caminho até chegar ao <recipiente>. Os 14 verbos apontados como os mais frequentes em cada construção são listados abaixo (em ordem decrescente dos mais frequentes para os menos frequentes):

Construção ditransitiva: give, tell, show, offer, cost, teach, wish, ask, promise, deny, award, grant, cause, drop.

Construção ditransitiva preposicionada: bring, play, take, pass, make, sell, do, supply, read, hand, feed, leave, keep, pay.

Os trabalhos supracitados mostram que o processo de decisão inconsciente através do qual falantes escolhem uma das duas construções dativas alternantes é influenciado por diversos fatores. Para propor um modelo que levasse em consideração toda essa complexidade, Bresnan et al. (2007) consideraram simultaneamente as seguintes variáveis para prever a seleção de uma das construções da AD:

Propriedades dos argumentos <recipiente> e <tema>: acessibilidade discursiva, extensão morfofonológica, animacidade, definição, pronominalidade, número (singular/plural), pessoa e concretude.

Paralelismo estrutural: refere-se à existência do mesmo tipo de estrutura no mesmo diálogo.

Sentidos específicos dos verbos: transferência metafórica (*give it some thought*), transferência de posse (*give you a book*), transferência futura de posse (*promise you a rose*); prevenção de posse (*deny you a day off*) e comunicação (*tell you a secret*).

O *corpus* analisado foi o SWITCHBOARD, composto por conversas telefônicas que abrangem as principais variedades do inglês americano, do qual foram extraídas 2.360 ocorrências de construções dativas, 1859 ditransitivas e 501 ditransitivas preposicionadas.

Para controlar múltiplas variáveis relacionadas a uma resposta binária (escolha de uma entre as duas construções alternantes), foi utilizada uma técnica estatística chamada de Regressão Logística, por meio da qual coeficientes foram atribuídos a cada variável supracitada, indicando a importância e a direção das mesmas em relação às construções dativas. Coeficientes positivos e negativos indicaram as variáveis que favorecem o uso da construção ditransitiva preposicionada e da construção ditransitiva, respectivamente. A interpretação dos resultados obtidos é resumida no Quadro 8.

Quadro 8 - Variáveis que favorecem a seleção de cada construção dativa com base na interpretação dos coeficientes obtidos através do teste de Regressão Logística por Bresnan et al. (2007)

Variáveis que favorecem a seleção da construção ditransitiva preposicionada	Variáveis que favorecem a seleção da construção ditransitiva
(a) acessibilidade do recipiente = não-dado (b) pronominalidade do recipiente = não-pronome (c) definitude do recipiente = indefinido (d) animacidade do recipiente = inanimado pessoa do recipiente = não-local número do tema = plural (e) extensão do recipiente = maior do que extensão do tema	(a) acessibilidade do tema = não-dado (b) pronominalidade do tema = não-pronome (c) definitude do recipiente = indefinido (d) número do recipiente = plural concretude do tema = não-concreto (e) extensão do tema = maior do que extensão do recipiente (f) paralelismo estrutural (existência da mesma estrutura no mesmo diálogo)

Os autores afirmam que as análises feitas indicam que a influência das variáveis mencionadas no Quadro 8 é significativa independentemente dos significados dos verbos. Destaque é dado às variáveis animacidade, pronominalidade e acessibilidade discursiva.

Uma possível limitação em relação ao modelo supracitado é, porém, apontada por Theijssse (2008). A autora questiona até que ponto um modelo construído apenas a partir da análise de conversas espontâneas sobre um único tópico e de artigos financeiros de jornais é capaz de prever o uso das construções dativas em outros gêneros orais e escritos. Para abordar essa questão, Theijssse aplica o modelo proposto por Bresnan et al. (2007) para a análise de dados do *corpus* ICE-GB, composto por uma gama maior de gêneros textuais. Apesar de os coeficientes obtidos para as variáveis terem indicado a direção esperada (na ditransitiva preposicionada, o tema é predominantemente pronominal e o recipiente é indefinido e não-local – terceira pessoa; na ditransitiva o tema é mais longo do que o recipiente, o recipiente é pronominal e o tema indefinido), o poder de previsão para os dados do ICE-GB foi consideravelmente menor do que o alcançado na análise do SWITBOARD. Com base nesse resultado, a autora sugere que gêneros textuais devem ser incluídos como uma variável adicional em pesquisas futuras sobre alternâncias sintáticas, pois diferenças entre tipos de textos podem afetar a previsibilidade do uso das construções ditransitivas. É importante destacar que os resultados de Theijssse reforçam a hipótese probabilística da língua.

O estudo de Theijssse (2008), além de explorar a relação entre variáveis linguísticas de

diversos níveis que poderiam influenciar a escolha entre as construções dativas, colocou em evidência como fatores extralinguísticos podem influenciar o uso de construções linguísticas. Outros trabalhos que investigam a alternância dativa do inglês à luz de fatores extralinguísticos através de análises de corpora incluem os estudos de Bresnan e Hay (2008), Kendall et al. (2011) e Theijssen et al. (s/d), nos quais variáveis sociolinguísticas como idade, gênero e variedade linguística são consideradas. Estudos como esses, ao invés de proporem regras determinísticas que apontam o que é gramatical ou não em uma língua, geralmente adotam teorias probabilísticas da língua, defendendo que o uso de estruturas sintáticas é influenciado por fatores linguísticos e extra-linguísticos cujas importâncias relativas podem variar entre diferentes grupos de falantes (BOD et al. 2003). De fato, os resultados alcançados nesses estudos corroboram tal proposta. Se, por um lado, tendências universais da AD têm sido cada vez mais evidenciadas (objetos animados usualmente precedem objetos inanimados, definitude precede indefinitude, constituinte menos extensos precedem constituintes mais extensos, pronomes precedem não pronomes), por outro lado, diferenças distribucionais sutis entre as duas construções dativas e entre o grau de importância das variáveis já mencionadas que influenciam a seleção das construções têm sido pouco a pouco identificadas em falantes que pertencem a grupos diferentes em termos de idade, gênero e variedade do inglês. Isso corrobora a noção de “restrições” revista pela Linguística Probabilística, como expomos na seção 2.2 desta tese.

Diferentes associações entre as construções dativas do inglês e fatores extra-linguísticos são apontadas também em estudos que utilizam um paradigma experimental. Bresnan e Ford (2010), por exemplo, constataram que, na tarefa experimental por eles utilizada, os homens australianos tenderam a produzir mais construções dativas preposicionadas do que mulheres australianas. Outros resultados interessantes que apontam nuances a respeito do uso da AD em diferentes variedades do inglês são mostrados abaixo:

- a) A variável animacidade exerce uma influência maior na AD na variedade falada do inglês da Nova Zelândia do que na variedade falada do inglês norte-americano (BRESNAN & Hay, 2008).
- b) Há mais argumentos <tema> e <recipiente> expressos por pronomes na variedade afro-americana do que em outras variedades do inglês (KENDALL et al., 2011).
- c) Americanos atribuem valores de naturalidade mais altos para construções ditransitivas preposicionadas do que britânicos e australianos (THEIJSEN et al., s/d).

- e) Americanos mais jovens atribuem um grau de naturalidade maior para construções ditransitivas preposicionadas do que americanos mais velhos (Ibidem).
- f) Britânicos mais jovens atribuem um grau de naturalidade maior para construções ditransitivas preposicionadas nas quais o <tema> é expresso através de um pronome do que britânicos mais velhos (Ibidem).
- h) Falantes do inglês australiano mostram-se mais sensíveis à influência da extensão do argumento <tema> em relação ao argumento <recipiente> do que falantes do inglês americano (Ibidem)

Outro dado de variabilidade em relação às construções dativas do inglês que devemos mencionar é a variação dialetal quanto às opiniões de falantes nativos de inglês sobre a possibilidade de certos verbos participarem ou não da alternância dativa, ou seja, de interagirem tanto com a construção ditransitiva quanto com a ditransitiva preposicionada (GRIES, 2003, p. 8). Juntamente com os resultados apresentados nessa seção, observações como essa são extremamente importantes para discussões sobre a aquisição de inglês/L2.

Os resultados apresentados nesta seção sugerem que a escolha inconsciente feita por falantes nativos de inglês entre a construção ditransitiva e a construção ditransitiva preposicionada em um contexto de uso é influenciada por fatores sintáticos, semânticos, discursivos e extra-linguísticos. Além disso, os trabalhos revisados ilustram como a disponibilidade de grandes corpora e a utilização de técnicas estatísticas modernas podem levar a uma compreensão muito mais refinada sobre fenômenos linguísticos que não poderia ser alcançada de outra forma.

Existe ainda um fator que pode influenciar a escolha das construções dativas do inglês que não foi explorado em nenhum dos estudos supracitados, mas que merece menção. Um dos papéis principais de uma construção de nível sentencial é focalizar a atenção do ouvinte/receptor em algum aspecto de uma cena (TOMASELLO, 2003a, p. 146). Segundo Goldberg (2006, p. 138), existe um consenso de que, na construção ditransitiva, o argumento <tema> representa a informação que se quer enfatizar, e que, na construção ditransitiva preposicionada, é o argumento <recipiente> que representa tal informação. Assim, a escolha entre as duas construções dativas no inglês pode ser feita também de acordo com qual informação o falante quer enfatizar, o que torna

a investigação sobre o uso dessas construções ainda mais complexa.

O Quadro 9 sintetiza as principais características de alguns estudos citados nesta seção.

Quadro 9 - Síntese dos estudos citados nesta seção sobre a alternância dativa do inglês

Estudo	Dados analisados	Variáveis consideradas
Gries (2003)	<i>Corpus</i> BNC	Tipo de processo descrito pela sentença; animacidade, extensão, grau de definição, pronominalidade e acessibilidade discursiva dos argumentos <tema> e <recipiente>.
Gries & Stefanowitsch (2004)	<i>Corpus</i> ICE-GB	Verbo utilizado.
Bresnan et al. (2007)	<i>Corpus</i> SWITCHBOARD e <i>Corpus</i> Wall Street Jornal.	Acessibilidade discursiva, extensão morfo-fonológica, animacidade, definição, pronominalidade, número (singular/plural), pessoa e concretude dos argumentos <recipiente> e <tema>; paralelismo estrutural; sentidos específicos dos verbos.
Theijssse (2008)	<i>Corpus</i> ICE-GB	Mesmas variáveis presentes em Bresnan et al. (2007); tipo de oração: principal, subordinada ou relativa; modo da sentença: declarativa, interrogativa ou imperativa; ordem das palavras: unmarked ou fronting; presença ou ausência de advérbio entre tema e recipiente.
Bresnan & Hay, 2008	<i>Corpus</i> ONZE (Origins of New Zealand English); Switchboard <i>Corpus</i> ; Treebank Wall-Street Journal	Complexidade sintática, pronominalidade, acessibilidade discursiva e animacidade dos argumentos <recipiente> e <tema> do verbo give; classe semântica do uso do verbo; variedade do inglês (americana e nova-zelandês)
Kendall al. (2011)	<i>Corpus</i> Switchboard e <i>Corpus</i> Wall Street Jornal. Dados orais do <i>Sociolinguistic Archive and Analysis Project</i> e de transcrições de entrevistas socio-linguísticas; cartas escritas do <i>Ottawa Repository of Early African American Correspondence</i> .	Mesmas variáveis presentes em Brenas e Hay (2008), porém as variedades do inglês foram as americana e afro-americana.
Theijsee et al. (s/d)	<i>Corpus</i> SWITCHBOARD; <i>corpus</i> ICE-GB; dados de julgamento de naturalidade.	Mesmas variáveis presentes em Bresnan et al. (2007); idade, gênero e variedade do inglês (britânica, americana e nova- zelandês)

Vimos nesta seção que múltiplas variáveis influenciam a chamada “alternância dativa” do inglês, e que a influência dessas variáveis apresenta probabilidades que podem variar segundo a região, idade e sexo de grupos de falantes de inglês/L1. Considerando essa realidade, passemos para a revisão de estudos sobre o uso das construções dativas em inglês/L2.

2.5.2 Estudos anteriores sobre construções dativas na interlíngua

Estudos sobre a alternância dativa na Interlíngua podem ser divididos em quatro grupos: (I) estudos que investigam a ordem de aquisição das construções dativas, (II) estudos que buscam identificar e compreender “violações” aos polos formais das construções dativas em produções da Interlíngua, (III) estudos que avaliam o conhecimento de aprendizes acerca de quais verbos participam ou não da alternância, e (IV) estudos que questionam se aprendizes são sensíveis aos fatores que influenciam a escolha das construções dativas em contextos de uso. Alguns desses diferentes tipos de estudos que, pautando-se em diferentes quadros teóricos, adotando diferentes metodologias e analisando dados de falantes nativos de diferentes L1s, contribuem para a discussão sobre a aquisição da estrutura argumental da L2 são revisados a seguir.

2.5.2.1 Ordem de aquisição das construções dativas

A ordem de aquisição das construções dativas tem sido discutida em muitos estudos. Mazurkewich (1984), por exemplo, pautando-se no gerativismo Chomskyano, investigou a ordem de aquisição das construções dativas do inglês por falantes nativos do francês e do inuktitut. Em uma tarefa de julgamento de gramaticalidade, 45 falantes nativos do francês (23 iniciantes, 7 intermediários e 15 avançados) e 38 falantes nativos de inuktitut (12 iniciantes, 8 intermediários e 18 avançados) decidiram se sentenças com verbos que são tradicionalmente considerados como participantes da alternância dativa (*give, read, send, throw, buy, make*) ou não (*explain, report,*

suggest, capture) representavam “um bom inglês”. Os resultados obtidos indicaram que aprendizes de ambas as L1s haviam adquirido a construção ditransitiva preposicionada antes da construção ditransitiva. Segundo a autora, essa seria a ordem de aquisição esperada para aprendizes de inglês/L2 independentemente de suas L1s porque a construção ditransitiva preposicionada seria uma estrutura não-marcada, parte da gramática central (princípios universais a todas as línguas), enquanto que a construção ditransitiva seria uma estrutura marcada, parte das regras específicas de cada língua.

Ao invés de recorrerem à teoria de marcação, alguns autores defendem que uma explicação mais plausível para a sequência de aquisição das construções dativas seria a influência translinguística. Zara (2009), por exemplo, através de uma tarefa de julgamento de aceitabilidade, concluiu que aprendizes brasileiros de inglês adquirem a construção ditransitiva preposicionada, existente em sua L1, primeiro que a construção ditransitiva, ausente na variedade padrão de sua L1. Seguindo Kellerman (1979), a autora sugere que a influência da L1 é suficiente para explicar seus resultados.

Outros dados relevantes para a discussão sobre a motivação da ordem de aquisição das construções dativas foram obtidos por Hamdan (1994), também através de tarefa de julgamento de gramaticalidade. Ao investigar a aquisição da alternância dativa da língua árabe moderna padrão por falantes nativos do inglês, a autora encontrou evidências da transferência de formas da L1 consideradas “marcadas” para a interlíngua, concluindo que uma teoria de marcação baseada em modelo de Gramática Universal não daria conta de explicar seus dados. Hamdan (Ibidem, p. 807) explica que tanto a língua árabe quanto a inglesa possuem verbos que participam da alternância dativa e que possuem significados bastante semelhantes, mas que as duas línguas se diferenciam em relação a quais verbos participam ou não da alternância. Por exemplo, o verbo correspondente a *sell* em árabe participa da alternância, enquanto que o verbo árabe correspondente a *buy* não. Através dos julgamentos de gramaticalidade de dez falantes nativos do inglês que passaram em um teste de proficiência em árabe moderno, foi constatado que a sequência de aquisição prevista segundo a teoria de marcação se aplica apenas para os verbos que não participam da alternância na L1 dos aprendizes. Por outro lado, ambas as construções dativas foram julgadas igualmente gramaticais se os verbos nelas utilizados também participavam da alternância na L1 dos aprendizes. Assim, a influência translinguística torna-se mais evidente.

Dados de produção que corroboram propostas de que a L1 exerce um papel importante na

ordem de aquisição de construções da L2 são fornecidos por Kang (2011). Nesse estudo, foram analisadas amostras do *corpus* KELC (KNU English Learner *Corpus*), composto por redações escritas por 315 crianças coreanas aprendizes de inglês como língua estrangeira classificadas em seis níveis de proficiência da língua-alvo. Foi constatado que apenas aprendizes dos grupos de proficiência avançado e pós-avançado produziram a construção ditransitiva do inglês e, assim, concluiu-se que a construção ditransitiva preposicionada do inglês é adquirida antes da construção ditransitiva do inglês por crianças coreanas. Segundo a autora, a maioria dos verbos dativos do coreano é usada apenas na construção ditransitiva preposicionada, o que explicaria a ordem de aquisição encontrada em seu estudo.

Outras sequências desenvolvimentais podem ser investigadas para se verificar a influência translinguística. Oh (2006), por exemplo, estudou a ordem de aquisição do tipo de recipiente, <alvo> ou <beneficiário>, da construção ditransitiva do inglês por falantes nativos do coreano. O autor concluiu que os sujeitos participantes de sua pesquisa tiveram mais facilidade em adquirir a construção ditransitiva do inglês com um recipiente <alvo>, como em “*Mary sent John a letter*”, do que a construção ditransitiva com um recipiente <beneficiário>, como em “*Mary baked John a cake*”, porque apenas uma estrutura semelhante à primeira se faz presente em sua L1.

Além de influenciar a ordem de aquisição das construções dativas da L2, como sugerem as pesquisas supracitadas, a L1 pode também influenciar produções da interlíngua que se distanciam daquelas encontradas na língua-alvo, como no seguinte exemplo extraído do Br-ICLE: “*We must give for them a second chance*”. Vejamos na próxima seção o que alguns estudos revelam nesse sentido.

2.5.2.2 Padrões sintáticos não esperados na língua-alvo

Resultados de pesquisas com aprendizes de inglês/L2 mostram que a interlíngua pode licenciar estruturas para expressar eventos de transferência de posse que não são esperadas segundo o que se observa na língua-alvo. Kang (2011), por exemplo, encontrou ocorrências de sentenças como “*I give to him a book*” e “*I give a book him*” em produções escritas de crianças coreanas, cuja língua materna, diferentemente do inglês, permite ordem livre dos argumentos

<recipiente> e <tema>. A autora explica que o caso dativo é atribuído pela adição de um morfema em coreano e que a ordem das palavras é menos importante, de forma que todas as sentenças seguintes são possíveis em coreano: *John gives a book to Mary* / *John gives to Mary a book* / *John gives Mary a book* / *John gives a book Mary*. Isso sugere que aprendizes coreanos de inglês transferem a ordem livre de sua L1 para a sua interlíngua. Semelhantemente, Valenzuela e Rojo (2008), ao analisarem dados de falantes nativos do espanhol extraídos do *International Corpus of Learner English*⁷ (ICLE), também encontraram sentenças do tipo **Peter gave to his father a present* / **Peter gave to David a present* / **Peter gave to him a present* nas produções de aprendizes.

Dados de julgamento de aceitabilidade também sugerem influência da estrutura argumental da L1 na aquisição da L2. Zeddari (2009), por exemplo, ao investigar a aquisição da alternância dativa do inglês por falantes nativos do árabe marroquino (33 aprendizes do nível pré-intermediário, 33 do nível intermediário e 33 do nível avançado), constatou que a ordem <Sujeito Verbo Oblíquo Objeto>, possível na L1 dos aprendizes, foi aceita em maior proporção por aprendizes do nível intermediário (Ibidem, p. 185), mas que efeitos de transferência da ordem livre dos argumentos <recipiente> e <tema> da L1 para a interlíngua podem ser encontrados até mesmo no nível avançado de proficiência (Ibidem, p. 267). Em estudo que revela resultados semelhantes a esses para a aquisição do inglês por falantes nativos do português do Brasil, Zara (2009) incluiu 30 sentenças em inglês com a ordem <Sujeito Verbo Oblíquo Objeto>, como em “*Mary gave to John a book* / *Maria deu para João um livro*”, em sua tarefa experimental e constatou que o grupo de aprendizes iniciantes (25 sujeitos) exibiu um comportamento estatisticamente diferente de falantes nativos do inglês em relação a 29 destas sentenças, tendendo a aceitá-las mais do que as construções ditransitivas do inglês presentes no experimento. No grupo de aprendizes intermediário (22 sujeitos), houve ainda diferença estatística entre o grupo experimental e o grupo controle em relação a 15 sentenças com a ordem <Sujeito Verbo Oblíquo Objeto>.

É fascinante pensar que aprendemos algo tão complexo quanto uma língua. Além de aprendermos os padrões sintáticos disponíveis em uma língua para expressar certos eventos, como os de transferência de posse, precisamos também saber quais verbos interagem com cada um desses padrões. Na próxima seção mostramos como aprendizes têm se saído nessa tarefa ao

⁷ <http://www.uclouvain.be/en-cecl-icle.html>

adquirirem as construções dativas do inglês.

2.5.2.3 Conhecimento sobre quais verbos participam da alternância dativa do inglês

O conhecimento que aprendizes possuem sobre quais verbos participam ou não da alternância dativa do inglês também tem sido foco de interesse de estudos sobre aquisição de inglês/L2 (DAVIES, 1994; INAGAKI, 1997; SOUSA, 2006; WHITE, 2003; WHONG-BARR & SCHWARTZ, 2002; ZARA, 2009; ZEDDARI, 2009). Relatamos abaixo o que alguns autores concluíram sobre essa questão. Antes disso, porém, chamamos a atenção para o fato de que, apesar de a possibilidade de verbos interagirem com as duas construções dativas do inglês ser tradicionalmente colocada de forma categórica (ex: alguns verbos alternam, outros não), assumimos que a interação entre verbos e as construções dativas do inglês representam associações probabilísticas influenciadas por fatores linguísticos e não linguísticos (ver seções 2.2 e 2.5.1 desta tese). Assim, quando relatamos, nas palavras dos autores, resultados sobre verbos que alternam ou que não alternam, entendemos, à luz da Linguística Probabilística, que alguns verbos frequentemente são encontrados em ambas as construções dativas do inglês, enquanto outros apresentam grande probabilidade de ocorrerem apenas em uma delas. Tendo esclarecido essa diferença de interpretação, passemos para os estudos revisados.

Inagaki (1997), seguindo a classificação semântica de verbos proposta por Pinker (1989), testou 32 falantes nativos do japonês e 32 falantes nativos do chinês com alto nível de proficiência em inglês em relação aos seus conhecimentos sobre a possibilidade de verbos de “causa instantânea de movimento balístico” (Throw-class), de “causa contínua de modo de movimento” (Push-class), “de comunicação” (Tell-class) e “de modo de fala” (Whisper-class) participarem da alternância dativa do inglês. O autor utilizou uma tarefa de julgamento de gramaticalidade e constatou que os sujeitos de sua pesquisa demonstraram o conhecimento testado apenas em relação a algumas classes semânticas de verbos. Ele relata que ambos os grupos de aprendizes foram capazes de diferenciar os verbos de “comunicação”, apontados pela literatura como participantes da alternância, dos verbos de “modo de fala”, apontados como não participantes da alternância. Contudo, segundo o autor, nenhum grupo alcançou um

conhecimento semelhante ao de falantes nativos do inglês em relação à não possibilidade prevista de os verbos de “causa contínua de modo de movimento” participarem da alternância. Similarmente, Zeddari (2009) aponta supergeneralizações na interlíngua árabe marroquino/inglês de aprendizes do nível avançado na língua-alvo indicadas pela sua grande aceitação dos verbos *whisper*, *push* e *donate* na construção ditransitiva, quando eles são previstos apenas na ditransitiva preposicionada.

O conhecimento sobre quais verbos participam da alternância dativa do inglês por falantes nativos do português do Brasil foi investigado por Sousa (2006) e por Zara (2009). Ambos os autores utilizaram tarefas de julgamento de aceitabilidade, porém a pesquisa de Sousa (2006) baseou-se no gerativismo Chomskyano e a de Zara (2009) fundamentou-se em abordagens construcionistas (CROFT & CRUSE, 2004; GOLDBERG, 1995; TOMASELLO, 2006). Os resultados obtidos nos dois estudos são relatados e comparados abaixo.

Sousa concluiu que brasileiros aprendizes do nível intermediário de proficiência em inglês não possuem conhecimento sobre a construção ditransitiva do inglês com base na tendência de esses aprendizes não aceitarem as sentenças ditransitivas gramaticais apresentadas na tarefa experimental de seu estudo, nas quais foram utilizados os seguintes verbos: *give*, *send*, *show*, *throw*, *get*, *bring*, *make*, *buy*, *find* e *draw*. Similarmente, Zara (2009) encontrou uma tendência à rejeição da construção ditransitiva do inglês pelos sujeitos de sua pesquisa classificados como aprendizes intermediários de inglês. Porém, a autora constatou que esses aprendizes sabem que o verbo *tell*, não incluído no estudo de Sousa (2006), pode ser usado no padrão <Sujeito Verbo Objeto1 Objeto 2>. Através da aplicação do Teste Exato de Fisher, a autora mostra que a aceitação desse verbo na construção ditransitiva pelo grupo de aprendizes intermediários é estatisticamente semelhante à aceitação do grupo de falantes nativos do inglês. Assim, vemos como a escolha dos verbos a serem incluídos em uma tarefa experimental sobre construções da estrutura argumental pode influenciar os resultados alcançados.

Uma mudança na interlíngua dos brasileiros aprendizes de inglês do nível intermediário para o nível avançado de proficiência na língua-alvo é constatada tanto por Sousa quanto por Zara. Sousa (2006) verificou que os aprendizes avançados apresentaram um comportamento semelhante aos dos falantes nativos do inglês na tarefa experimental em relação à construção ditransitiva com verbos mais comuns nesta construção, como *buy*, *show*, *bring*, *send*, *get* e *give*” (Ibidem, p. 63). O autor afirma, porém, que os aprendizes “tendem a não saber nada” sobre

sentenças com verbos menos comuns na construção ditransitiva, como *throw*, *draw* e *find*, (Ibidem). Zara (2009), por sua vez, constatou que brasileiros aprendizes avançados de inglês foram estatisticamente semelhantes aos falantes nativos da língua alvo em relação à aceitação de todas as sentenças apresentadas em seu experimento com verbos que ocorrem na construção ditransitiva (*tell*, *teach*, *give*, *hand*, *bring*, *build*, *make*, *buy*, *find*, *offer*, *promise*). No que tange às sentenças com verbos que não são tradicionalmente previstos na alternância dativa, Zara (2009) constatou supergeneralizações nas interlínguas de aprendizes avançados de inglês. Não obstante, a aplicação do Teste Exato de Fisher revelou que houve semelhanças estatísticas entre as respostas dos aprendizes e dos falantes nativos em relação à maioria das sentenças apresentadas na tarefa experimental com verbos não previstos na construção ditransitiva (*donate*, *construct*, *purchase*, *present*, *compose*, *collect*, *whisper*, *pick*, *choose*). Isso indica que, pelo menos em contextos de tarefas de julgamento de gramaticalidade, a aceitação de verbos nas construções dativas por falantes nativos do inglês é tão flexível quanto a de aprendizes do nível avançado em inglês.

Ao compararmos resultados de estudos, como fizemos acima, precisamos levar em consideração diferenças metodológicas entre eles. Enquanto Zara classificou os aprendizes participantes de sua pesquisa como iniciantes, intermediários ou avançados segundo seu desempenho em uma versão impressa do teste de nivelamento online da Cambridge University, Sousa considerou aprendizes e professores de inglês de diferentes instituições de ensino como sendo dos níveis intermediário e avançado, respectivamente, sem utilizar nenhuma medida de nível de proficiência. Além disso, há diferenças regionais e dialetais entre os sujeitos das pesquisas. Zara coletou dados da interlíngua no estado de Minas Gerais e seu grupo controle era composto por americanos, australianos e britânicos. Sousa, por sua vez, realizou seu estudo no estado de São Paulo e seu grupo controle foi formado apenas por falantes do inglês americano. Diferenças como as apontadas entre os estudos de Sousa e de Zara podem explicar porque dois estudos distintos investigando a mesma questão podem alcançar diferentes resultados.

Os estudos revisados acima sugerem que aprendizes demonstram alguns conhecimentos sobre quais verbos podem interagir com as construções dativas do inglês. A próxima pergunta a se fazer é: Em contextos reais de uso, aprendizes sabem como selecionar uma das construções quando as duas são possíveis? Alguns estudos que abordam essa questão são revisados na seção 2.5.2.4.

2.5.2.4 Sensibilidade de aprendizes aos fatores que influenciam a escolha por uma das construções dativas

Descobertas recentes de pesquisas sobre a alternância dativa na L1, tais como a observação de violações a restrições rígidas postuladas para essa alternância em dados de uso real do inglês, têm levado alguns pesquisadores a mudarem seu foco de interesse em relação à investigação sobre a alternância dativa do inglês na interlíngua (WOLK et al. 2011, p. 1). Ainda em pequeno número, pesquisadores têm investigado se aprendizes de inglês/L2 assemelham-se aos falantes nativos da língua-alvo no que tange à sua sensibilidade a diversos fatores envolvidos na escolha entre as duas construções dativas do inglês (CALLIES & SZCZESNIAK, 2008; FRISHKOFF et al., 2008; GONZÁLEZ-ROMERO, 2011; GRIES & WULFF, 2005; PARK, 2011; VALENZUELA & ROJO, 2008; WOLK et al., 2011; WURM, KONIECZNY & HEMFORTH; 2012; ZEDDARI, 2009). Vejamos o que alguns deles têm revelado.

Influências de efeitos de *priming*, isto é, da tendência para a repetição de estruturas sintáticas usadas anteriormente, bem como a influência de verbos específicos na seleção das construções dativas do inglês/L2 já foram identificadas. Gries e Wulff (2005), por exemplo, analisaram as respostas dadas em uma tarefa de completar sentenças realizada por 64 falantes nativos do alemão, alunos do curso de graduação em inglês na Universidade de Hamburg, e encontraram efeitos de *priming* assim como esperado para falantes nativos da língua-alvo. Além disso, os autores observaram que os aprendizes eram sensíveis aos verbos que exibem preferência por uma das construções dativas do inglês. Tal conclusão baseia-se no fato de que os verbos fortemente associados à construção ditransitiva, como *give* e *show*, foram mais propícios a causar *priming* dessa construção, sendo o mesmo verdadeiro para verbos como *post* and *sell* em relação à construção ditransitiva preposicionada (Ibidem, 188).

Evidências de que aprendizes de inglês/L2 falantes nativos do alemão são capazes de perceber tendências lexicais estatísticas em relação às construções dativas são também fornecidas por Wolk et al. (2011). Por meio de experimentos de processamento da língua falada e escrita, os autores constataram, dentre outras coisas, que quanto mais a combinação entre verbo e construção testemunhada pelos aprendizes durante o experimento fosse próxima da esperada, mais rápida era a leitura da sentença.

Dados empíricos mostram também que aprendizes são sensíveis à influência da acessibilidade discursiva do recipiente e do tema na escolha por uma das construções dativas do inglês. Por exemplo, Park (2011) coletou dados de 23 coreanos aprendizes de inglês/L2 através de tarefas experimentais que objetivavam mostrar as escolhas sintáticas dos aprendizes em contextos específicos de uso. Os sujeitos da pesquisa foram recrutados na Universidade do Havaí, tendo alcançado mais de 70% de acerto em relação à morfossintaxe e ao léxico do inglês em uma tarefa de descrição de figura. Duas tarefas orais de preferência contextualizada foram utilizadas. Em uma delas, as informações dadas eram expressas através de NPs lexicais (tarefa NP lexical), na outra, através de pronomes (tarefa pronome). Em ambos os casos, os sujeitos ouviam pequenas narrativas acompanhadas de ilustrações e seguidas pela pergunta do narrador para os personagens da história: “*Hey, Bobo and Kayu, do you know what happened next?*”. Imediatamente após cada personagem fornecer uma resposta com uma das construções dativas, a pergunta “*Which one is the better way to say it?*” (“Qual é a melhor forma de dizer isso?”) era feita aos aprendizes. As sentenças do experimento apresentavam ora o <tema> como informação dada e o <recipiente> como informação nova, ora o <recipiente> como informação dada e o <tema> como informação <nova>. Seis verbos que participam da AD foram utilizados nas sentenças testes do experimento: *bring, give, mail, sell, send* e *show*. No geral, os aprendizes mostraram preferência pela ordem “dado antes do novo”.

A sensibilidade à generalização “informação dada antes da informação nova” aplicada à seleção das construções dativas do inglês/L2 foi também investigada por Callies e Szczesniak (2008), os quais incluíram em sua pesquisa a influência da extensão dos argumentos <recipiente> e <tema> na escolha por uma das construções dativas do inglês. Os autores fizeram um estudo baseado em *corpus*, no qual foram analisados dois subcorpora de produções escritas de falantes não-nativos e um subcorpus de falantes nativos de inglês. Os dados de aprendizes avançados de inglês/L2 – falantes nativos do alemão e do polonês – foram retirados do *International Corpus of Learner English* (ICLE). O subcorpus de produções escritas de falantes nativos da língua-alvo foi extraído do *corpus LOCNESS*, composto por textos produzidos por estudantes britânicos e americanos. A análise de Callies e Szczesniak focalizou 15 verbos dativos altamente frequentes. Os resultados encontrados nos subcorpora dos aprendizes foram semelhantes ao que foi constatado no subcorpus da L1: aprendizes colocam informações mais curtas e dadas discursivamente antes de informações novas e mais longas. Com base nesses resultados, os

autores concluem que as variáveis “acessibilidade discursiva” e “peso sintático” influenciam as escolhas das construções dativas por aprendizes de inglês/L2. Além disso, é interessante ressaltar que, corroborando os resultados de estudos experimentais já citados nessa seção, Callies e Szczesniak constataram que a preferência pelo uso de uma das construções alternantes com um verbo específico exibida em corpora de falantes nativos do inglês é ecoada nos corpora de aprendizes. Por exemplo, o verbo *tell* é usado predominantemente na construção ditransitiva nos três corpora analisados (Ibidem, p. 177).

Dados do ICLE foram analisados também em outras pesquisas que buscaram compreender a variação de uso das construções dativas do inglês por falantes nativos de outras línguas. González-Romero (2011), por exemplo, comparou fatores que influenciam a escolha dessas construções no componente espanhol do ICLE e no subcorpus *British Academic Written English* do *British National Corpus*. Segundo a autora, os resultados obtidos sugerem que tanto os falantes nativos quanto os não-nativos de inglês são influenciados pelos mesmos fatores, a dizer, o contraste entre informação dada e nova, o grau de definição dos NPs, a pronominalidade, a animacidade e a extensão dos NPs. Não obstante, em comparação aos corpora de falantes nativos da língua-alvo, foi constatada uma superprodução da construção dativa preposicionada no corpus de aprendizes, o que seria, segundo a autora, resultado da influência da L1 na interlíngua.

Com base nos resultados dos quatro tipos de estudos sobre a alternância dativa do inglês na interlíngua discutidos separadamente acima, podemos dizer que há evidências de que: (I) a L1 dos aprendizes influencia a ordem de aquisição das construções dativas da L2 e algumas produções da interlíngua que se distanciam de formas esperadas para a língua-alvo, (II) aprendizes em níveis mais avançados de proficiência na L2 demonstram conhecimentos semelhantes aos de falantes nativos da língua-alvo no que diz respeito à categorização de verbos que participam ou não da alternância dativa e à sensibilidade a fatores que influenciam a escolha por uma das construções alternantes. O Quadro 10 mostra uma síntese dos estudos revisados nesta seção.

Quadro 10 - Síntese dos estudos citados sobre a alternância dativa na interlíngua

Estudo	L1 dos aprendizes	Metodologia
Mazurkewich (1984)	Francês e Inuktitut	Julgamento de gramaticalidade
Inagaki (1997)	Japonês e chinês	Julgamento de gramaticalidade
Gries e Wulff (2005)	Alemão	Tarefa de completar sentenças
Sousa (2006)	Português do Brasil	Julgamento de gramaticalidade
Callies e Szczesniak (2008)	Alemão e polonês	Análise de <i>corpus</i>
Valenzuela e Rojo (2008)	Espanhol	Análise de <i>corpus</i>
Zara (2009)	Português do Brasil	Julgamento de aceitabilidade
Zeddari (2009)	Árabe marroquino	Julgamento de aceitabilidade
Kang (2011)	Coreano	Análise de <i>corpus</i>
Park (2011)	Coreano	Tarefas experimentais orais
González-Romero (2011)	Espanhol	Análise de <i>corpus</i>
Wolk et al. (2011)	Alemão	Experimentos psicolinguísticos

Uma vez que o presente trabalho investiga a interlíngua português(Brasil)/inglês, faz-se necessário conhecer o que pesquisas mostram sobre construções dativas do PB. É esse o assunto da próxima seção.

2.6 Estudos anteriores sobre construções dativas no português do Brasil

Na variedade culta do PB, eventos de transferência de posse podem ser expressos através de estruturas que variam em relação a qual preposição é utilizada (9), à ordem dos complementos

dos verbos (10) e à pronominalização com o clítico dativo ou na estrutura preposicionada (11) (BISPO & SALLES, 2005; GOMES, 2003;). Os exemplos abaixo, extraídos do *corpus* Humanas (ver seção 3.1), ilustram essas variações:

9) Alternância entre as preposições “a” e “para”:

- a) Como vocês respondem à crítica de certos filósofos de que o método é uma vulgarização da filosofia, que é impossível **ensinar filosofia a** crianças?
- b) Também ficou oito anos em volta redonda, ensinando música **para** crianças carentes.

10) Alternância entre as ordens dos argumentos:

- a) Dê autonomia total <tema> **aos aprendizes de professor** <recipiente>.
- b) A nova política será rapidamente implementada e dará **a universidades e instituições de pesquisa** <recipiente> maior autonomia <tema>.

11) Alternância entre a pronominalização com o clítico dativo e na estrutura preposicionada:

- a) Perguntou-me por meu pai e eu **lhe** dei informações. / O público identificou-se com o seu sofrimento e deu-**lhe** o prêmio.
- b) O delegado perguntou o que queria que fizesse com ele, eu falei: “não faz nada não, dá uns conselhos **para ele**”.

Um padrão frasal semelhante ao da construção ditransitiva do inglês já foi identificado em algumas variedades do português popular falado (12), como, por exemplo, na Zona da Mata Mineira (SCHER, 1996), no Rio de Janeiro (GOMES, 2003) e na Bahia (LUCCHESI & MELLO, 2009). O apagamento da preposição introdutória do dativo na ordem <Verbo Objeto_{Tema} Objeto_{Recipiente}> (13) também já é documentado (GOMES, 2003):

12) Maria deu Ø Pedro o livro.

13) Maria deu o livro Ø Pedro.

Apesar de haver estudos sobre as variações ilustradas nas sentenças acima (ARMELIN, 2010; BISPO, 2004; CAVALCANTE, 2009; ; FURTADO DA CUNHA, 2011; GOMES, 2003;

LUCCHESI & MELLO, 2009; SCHER, 1996), eles não se fundamentam em dados de corpora representativos do PB, como veremos abaixo.

Bispo (2004) conduziu entrevistas sociolinguísticas para investigar mudanças em andamento na sintaxe do dativo no português falado no Rio de Janeiro e na Paraíba. A autora focalizou a alternância entre as preposições introdutoras do dativo e entre os pronomes (no sintagma preposicionado *versus* na forma cliticizada). Os dados da pesquisa foram coletados com 2 falantes cariocas e 2 falantes paraibanos. As amostras obtidas indicam (I) o aumento do uso da preposição “a” no português da Paraíba (4 ocorrências de “a” *versus* 0 ocorrência de “para”), (II) o aumento do uso da preposição “para” no português do Rio de Janeiro (35 ocorrências de “para” *versus* 5 ocorrência de “a”) e (III) a diminuição do uso do clítico de 3ª pessoa em ambas as variedades do português estudadas. Em relação a (II), Bispo (2004) corrobora o estudo anterior de Gomes (2003), no qual a autora constatou, a partir da comparação entre dados coletados em dois períodos distintos, que “para” está substituindo “a” na língua falada no Rio de Janeiro.

Além da variação entre as preposições “para” e “a”, Gomes (Ibidem) investigou a omissão dessas preposições e a variação na ordem dos complementos de verbos bitransitivos. A autora analisou duas amostras de entrevistas sociolinguísticas que fazem parte do Programa de Estudos sobre o Uso da Língua: uma de 1980, com 64 informantes, e outra de 2000, com 32 informantes do português falado no Rio de Janeiro. Juntamente à popularização da preposição “para” na fala, como supracitado, as seguintes constatações foram feitas no que tange ao uso das preposições:

- I. A preposição “para” é mais utilizada (1) quando o complemento dativo é adjacente ao verbo, (2) quando o argumento <alvo> é humano e (3) com verbos de conteúdo lexical (*full content verbs*).
- II. A preposição “a” é mais frequentemente utilizada (1) distante do verbo, (2) em contextos formais e (3) por falantes com níveis mais altos de educação formal.
- III. A omissão de preposição é rara e ocorre apenas com verbos que permitem o uso da preposição “a”.

Em relação à ordem dos complementos, a ordem < Suj. Verb. Obj. Obl.> se mostrou a

configuração mais frequente para complementos de verbos bitransitivos. Não foi identificado, contudo, nenhum contexto no qual uma das variantes ocorreu “categoricamente”, como coloca a autora. Assim, não foi possível definir um contexto no qual apenas o uso de uma das variantes é permitido. Não obstante, a análise quantitativa dos dados apontou fatores sintáticos e semânticos que favorecem o uso de cada variante. A ordem <Suj. Verb. Obj. Obl.> é favorecida quando o verbo denota transferência material e quando o complemento dativo é consideravelmente maior (mais que 5 sílabas) do que o objeto direto e carrega informações novas. Já a ordem <Suj. Verb. Obl. Obj. > é mais frequente com verbos de transferência verbal ou perceptual.

Os padrões <Suj. Verb. Obl. Obj. > e <Suj. Verb. Obj.1 Obj.2> também foram investigados no português falado na Zona da Mata Mineira (PBM). Partindo do quadro teórico do Programa Minimalista de Chomsky (1993, 1995, 1995 *apud* SCHER, 1996, p. 106), Scher (Ibidem) objetivou propor uma representação sintática para a ordem <Suj. Verb. Obl. Obj.> e explicar a motivação para a ocorrência da estrutura <Suj. Verb. Obj.1 Obj.2>. Com base na análise de dados do PBM e à luz da teoria adotada, a autora argumenta que a ordem <Suj. Verb. Obl. Obj.> é motivada por um fator discursivo: quando o objeto indireto for o tópico do discurso, ele ocupará a posição imediatamente pós-verbal.

14) O que a Maria deu pro Pedro? A Maria deu pro Pedro um livro.

O padrão <Suj. Verb. Obj.1 Obj.2>, por sua vez, seria, segundo a autora, resultado do apagamento da preposição “a” na estrutura <Suj. Verb. Obl. Obj.>. A omissão da preposição ocorreria devido a fatores de natureza morfo-fonológica, tais como a reestruturação silábica. Scher ressalta, assim como vimos em Gomes (2003), que essa omissão é possível apenas com verbos que selecionam a preposição “a”.

Contar o caso (para/a) o público → contar o público o caso

Dar um presente (para/a) → dar o pai um presente

Cantar (uma canção) para os amigos → *cantar os amigos uma canção

Comprar um presente para a mãe → * comprar a mãe um presente

Scher defende que a omissão da preposição no PBM diferencia-se do fenômeno do inglês.

Primeiramente, no PBM a omissão acontece independentemente da posição em que o complemento dativo ocorre, ou seja, não é necessário haver adjacência entre o verbo e o dativo (15 – 16) (SCHER, 1996, p. 39). Além disso, o PBM, diferentemente do inglês, não licencia a passivização do dativo (17) (Ibidem, p. 28).

15) Mostra o carrinho os meninos!

16) Dá o recado o seu irmão!

17) *Os meninos foram dados um livro.

É interessante notar, contudo, que uma omissão da preposição introdutória do dativo mais próxima do fenômeno em inglês, ou seja, uma omissão que só ocorre em contexto de adjacência estrita entre o complemento e o verbo, foi constatada na variedade afro-brasileira do português de Helvécia, município de Nova Viçosa, interior da Bahia, por Lucchesi e Mello (2009). Os autores discutem a emergência desse padrão no português e defendem que, uma vez que a construção ditransitiva não se faz presente nas línguas românicas, mas é encontrada na maioria das línguas crioulas (Ibidem, p. 154), sua ocorrência nas variedades populares do português no Brasil relaciona-se historicamente com o contato do português com as línguas africanas. As estruturas dativas analisadas pelos autores foram extraídas de 48 entrevistas sociolinguísticas por eles conduzidas, totalizando 40 horas de gravação. Assim como constatado nos demais estudos citados nesta seção, a ocorrência da construção ditransitiva foi limitada: dentre as 358 estruturas analisadas, 83% representam construções com o objeto indireto introduzido por uma preposição (ela dá comida pas criança / escreveu pra nós uma carta) e apenas 17% são ocorrências da construção de objeto duplo (ele vendia compade Jacó porco gordo). Os fatores identificados como favorecendo a ocorrência da construção ditransitiva na amostra analisada foram:

1. Semântica dos verbos: a construção de objeto duplo foi mais frequente com verbos que implicam a transferência material ou perceptual entre duas partes (dar, pedir, doar, mostrar, encomendar, devolver, emprestar, vender, entregar, ceder, passar, pagar etc.). Já os verbos de comunicação verbal (dizer, falar, contar, perguntar, ensinar, agradecer, escrever, explicar, avisar, alertar etc.) foram os menos frequentes nessa construção.

2. Faixa etária: falantes mais velhos usaram a construção de objeto duplo com mais frequência do que falantes mais jovens, o que aponta para um declínio no uso da construção.

3. Nível de educação formal: a construção de objeto duplo ocorreu mais frequentemente entre os analfabetos, o que evidencia como a escola tem sido um meio de estabelecimento de normas do comportamento linguístico hegemônico nas comunidades periféricas.

Vimos nesta seção estudos que contribuem para a nossa compreensão sobre as construções dativas do português. Porém, ressaltamos que tais estudos ainda são limitados no que tange ao número de dados e de variedades linguísticas analisadas, bem como em relação a informações sobre probabilidades de ocorrências de construções alternantes. Portanto, é necessário que sejam feitas mais pesquisas sobre as construções dativas do português baseadas em corpora representativos.

O Quadro 11 resume os trabalhos sobre o português citados nesta seção.

Quadro 11 - Síntese dos estudos citados sobre a alternância dativa no português do Brasil

Estudo	Variedade do português falado	Objetivos
Scher (1996)	na Zona da Mata Mineira	Propor uma representação sintática para a ordem <Suj. Verb. Obl. Obj. > e explicar a motivação para a ocorrência da estrutura <Suj. Verb. Obj.1 Obj.2>.
Gomes (2003)	no Rio de Janeiro	Analisar a alternância entre as preposições e entre as ordens dos complementos de verbos bitransitivos; omissão da preposição introdutória do dativo.
Bispo (2004)	no Rio de Janeiro e na Paraíba	Analisar a alternância entre as preposições (para <i>versus</i> a) e entre os pronomes (no sintagma preposicionado <i>versus</i> na forma cliticizada).
Lucchesi e Mello (2009)	no município de Nova Viçosa, interior da Bahia	Discutir a emergência da construção ditransitiva os e fatores linguísticos e extra-linguísticos relacionados à sua ocorrência no português falado.

Neste capítulo apresentamos as fundamentações teórico-metodológicas e os resultados de estudos anteriores que serviram de base para esta tese. No próximo capítulo, descrevemos a metodologia utilizada para o desenvolvimento da nossa pesquisa.

3. METODOLOGIA

Para respondermos nossas perguntas de pesquisa, realizamos análises de corpora. Na análise dos corpora da interlíngua português/inglês, buscamos descrever como os aprendizes brasileiros de inglês representados nos corpora selecionados tendem a expressar eventos de transferência de posse em produções na língua-lavo. A partir do contraste entre os dados dos corpora dos aprendizes e os dados dos corpora de falantes nativos de inglês, identificamos em quais aspectos os aprendizes assemelham-se aos falantes nativos da língua-alvo em relação a diversos aspectos do uso das estruturas focalizadas. A análise dos corpora do português, por sua vez, permitiu que levantássemos hipóteses sobre a influência da L1 na interlíngua.

As seções que compõem este capítulo descrevem os corpora analisados (3.1), o *software* utilizado para a análise desses corpora (3.2), os critérios de seleção de expressões dos corpora do inglês e da interlíngua (3.3), as categorias para a análise dessas expressões (3.4), a análise dos corpora do português e as análises quantitativas dos dados (3.6).

3.1 Corpora analisados

Os dados da interlíngua português/inglês que mostraremos no próximo capítulo correspondem a produções orais e escritas de brasileiros aprendizes de inglês dos níveis intermediário ao avançado na língua-alvo retiradas de duas amostras de corpora da interlíngua português/inglês ainda em fase de compilação durante a realização desta pesquisa. As produções orais são de uma amostra do componente brasileiro do *The Louvain International Database of Spoken English Interlanguage* (LINDSEI), associado ao projeto *Corpus Oral Brasileiro de Aprendizes de Inglês*⁸. Já as produções escritas são de uma amostra do Br-ICLE, sub-corpus

⁸<http://www.lettras.ufmg.br/CMS/index.asp?pasta=leel&path=2010122910448.asp&title=Projetos%20em%20andamento>

brasileiro do *International Corpus of Learner English*⁹. Ambas as amostras de corpora de aprendizes¹⁰ são do tipo “não-etiquetadas” em relação a informações morfossintáticas.

O *corpus* LINDSEI—BR (MELLO et al., 2012) foi iniciado em 2011 e segue os padrões internacionais do LINDSEI. Os dados já coletados, porém ainda em fase de transcrição, consistem em padrões de fala quase-espontâneos produzidos por estudantes universitários, do nível intermediário ao avançado em inglês/L2, durante a realização de três tipos de tarefas: uma narrativa, uma discussão livre e uma descrição. A amostra do LINDSEI-BR analisada nesta pesquisa possui 30.952 palavras e representa 15 falantes.

O *corpus* Br-ICLE é composto por textos argumentativos escritos por estudantes universitários. Os tópicos dos textos são variados, incluindo “Love is an art”, “Should doctors help patients die?” e “University is necessary to get a good job now a days”. Quando completo, o *corpus* totalizará 200.000 palavras. A amostra analisada nesta pesquisa contém 159.364 palavras, distribuídas em 332 textos que correspondem a 332 aprendizes.

Para a descrição do inglês/L1, analisamos o *subcorpus* escrito (*Written*) e o *subcorpus* oral (*Spoken*) do componente britânico do *International Corpus of English* (ICE-GB), um *corpus* anotado com informações sobre classes gramaticais, estrutura sintática e traços morfológicos. O ICE-GB é considerado um *corpus* representativo do inglês britânico, tanto por possuir grande extensão, sendo composto por 200 textos escritos (423.702 palavras) e 300 textos falados (637.562 palavras) produzidos entre 1990 e 1993, quanto por incluir uma grande variedade de gêneros e registros, tais como diálogos particulares presenciais e por telefone, entrevistas transmitidas em rádios e televisão, debates parlamentares, cartas, textos acadêmicos e romances. Enfatizamos desde já que os resultados apresentados e discutidos sobre o inglês/L1 no próximo capítulo representam a variedade britânica expressa pelo ICE-GB, lembrando que as propriedades probabilísticas das construções dativas do inglês podem apresentar diferenças de uma variedade do inglês para outra.

A L1 dos falantes da interlíngua português/inglês foi analisada através de dois corpora: o C-ORAL-BRASIL, um *corpus* oral, e o Humanas, um *corpus* escrito. A inclusão de um *corpus* oral do português/L1 nesta pesquisa baseia-se na hipótese de que padrões da L1 frequentes na fala estão mais ativados nas mentes dos falantes e exibem, por isso, mais probabilidade de serem

⁹ <http://www.uclouvain.be/en-cecl-icle.html>

¹⁰ É importante destacar que “aprendizes brasileiros” nesta tese correspondem a estudantes universitários a partir do nível de proficiência intermediário em inglês/L2.

transferidos de forma involuntária para a interlíngua (POULISSE & BONGAERTS, 1994). A análise de um *corpus* escrito da L1, por sua vez, ao mesmo tempo em que pode revelar padrões que se diferem daqueles identificados na fala, foi importante nesta pesquisa por corresponder ao mesmo canal representado pelo *corpus* da interlíngua português/inglês com maior extensão que analisamos, o Br-ICLE.

O C-ORAL-BRASIL é um “*corpus* de fala espontânea do português brasileiro, representado pela diatopia mineira, majoritariamente proveniente da região metropolitana de Belo Horizonte” (MELLO, 2012, p. 50). O *corpus*, etiquetado com informações morfossintáticas, contém 208.130 palavras distribuídas em 139 textos que representam uma variedade de situações comunicativas em contextos privados e públicos. O Humanas, por sua vez, possui 1.786.289 palavras distribuídas em textos do domínio Ciências Humanas pertencentes aos gêneros informativo, científico, jurídico e institucional, distribuídos por diversos meios, dentre eles a internet, jornais, periódicos e revistas. O Humanas é um sub*corpus* do Lácio-Ref, um *corpus* não etiquetado de referência da norma culta do português brasileiro que possui atualmente 8.291.818 palavras e que faz parte do projeto Lácio-Web (ALUÍSIO et al., 2004). Esse projeto tem como objetivo disponibilizar corpora escritos para estudos linguísticos e para o público em geral, possuindo seis corpora compostos por textos escritos entre os anos de 1994 e 2003. Quatro categorias classificam os materiais que fazem parte do projeto: gênero (científico, de referência, informativo, jurídico, prosa, poesia, drama, institucional e técnico-administrativo), tipo de texto (apostila, artigo, carta, ensaio etc.), domínio (Ciências Agrárias, Ciências Humanas, Ciências Biológicas, Ciências Sociais Aplicadas, Ciências da Saúde, Ciências Exatas e da Terra, Religião & Pensamento e Generalidades) e meio de distribuição (CD-ROM, manuscrito, internet, panfleto etc.).

O Quadro 12 compara os corpora analisados em termos de extensão e variedade linguística representada. Os tamanhos relativos dos corpora são indicados segundo a classificação proposta por Sardinha (2000b, p. 7).

Quadro 12 - Resumo das características dos corpora analisados. Indicação dos tamanhos relativos dos corpora baseada na classificação de Sardinha (2000b, p.7).

Corpus	Número de palavras	Tamanho relativo	Variedade representada
LINDSEI-BR	30.952	Pequeno	Interlíngua português/inglês falado dos níveis intermediário ao avançado de proficiência em inglês.
Br-ICLE	159.364	Pequeno-médio	Interlíngua português/inglês escrito dos níveis intermediário ao avançado de proficiência em inglês.
ICE-GB <i>Spoken</i>	637.562	Médio	Inglês britânico falado.
ICE-GB <i>Written</i>	423.702	Médio	Inglês britânico escrito.
C-ORAL-BRASIL	208.130	Pequeno-médio	Português informal falado.
Humanas	1.786.289	Médio-grande	Português “culto” escrito.

Ainda comparando os corpora analisados nesta pesquisa, notamos que eles apresentam diferenças em termos dos períodos de produção dos textos: no *corpus* do inglês, as produções foram feitas entre 1990 e 1993, no *corpus* escrito do português, entre 1994 e 2003, no *corpus* oral do português, entre 2006 e 2011, e, nos corpora da interlíngua, elas ainda estavam sendo feitas no período de realização desta pesquisa. Embora reconheçamos que “toda a língua falada no mundo está em constante processo de mudança (GABAS JR, 2001, p. 81), baseamo-nos na afirmação de que uma variedade linguística pode ser estudada em intervalos de 25 a 30 anos ou mais (AYRES-BENNETT, 2004) para considerarmos que as diferenças entre os períodos de tempo representados nos corpora supracitados não representam um problema para a nossa pesquisa.

Os corpora descritos nesta seção foram analisados com a ajuda do *software* R, sobre o qual falamos a seguir.

3.2 Software que auxiliou a análise dos corpora

Grande parte da análise de *corpus* depende de ferramentas computacionais (SARDINHA,

2000a). Com tais ferramentas, podemos calcular o número de palavras de um *corpus*, gerar listas de frequência de palavras, extrair informações específicas de arquivos de *corpus* etc. Dentre os diversos programas disponíveis para a análise de corpora, escolhemos utilizar o R (R Core Team, 2013), um software que tem ganhado cada vez mais adeptos e colaboradores por apresentar várias vantagens (PETERNELLI & MELLO, 2011). Citaremos aqui três delas. Em primeiro lugar, o R é um software gratuito, que pode ser obtido na página do projeto. Além disso, o R é uma ferramenta muito versátil, permitindo em um só ambiente a manipulação de dados, a realização de cálculos e a geração de gráficos (GRIES, 2009a). Por último, devemos enfatizar que o R permite maior controle por parte do pesquisador sobre a análise que está sendo feita. Através de scripts do R, que correspondem a conjuntos de comandos que “dizem” ao R “o que fazer”, o linguista pode, por exemplo, definir como uma busca específica por dados é realizada considerando características específicas de um arquivo de texto.

Existem diversas publicações disponíveis para aqueles que desejam aprender sobre o R (BAAYEN, 2008; GRIES, 2009a; PETERNELLI & MELLO, 2011; ZARA, 2013). Introduzimos aqui algumas noções básicas para aqueles que ainda não conhecem o programa. Ao abrirmos o RStudio¹¹, visualizamos a janela exibida na Figura 1, na qual o símbolo “>” indica que o R está pronto para receber comandos.

¹¹ <http://www.rstudio.com/>

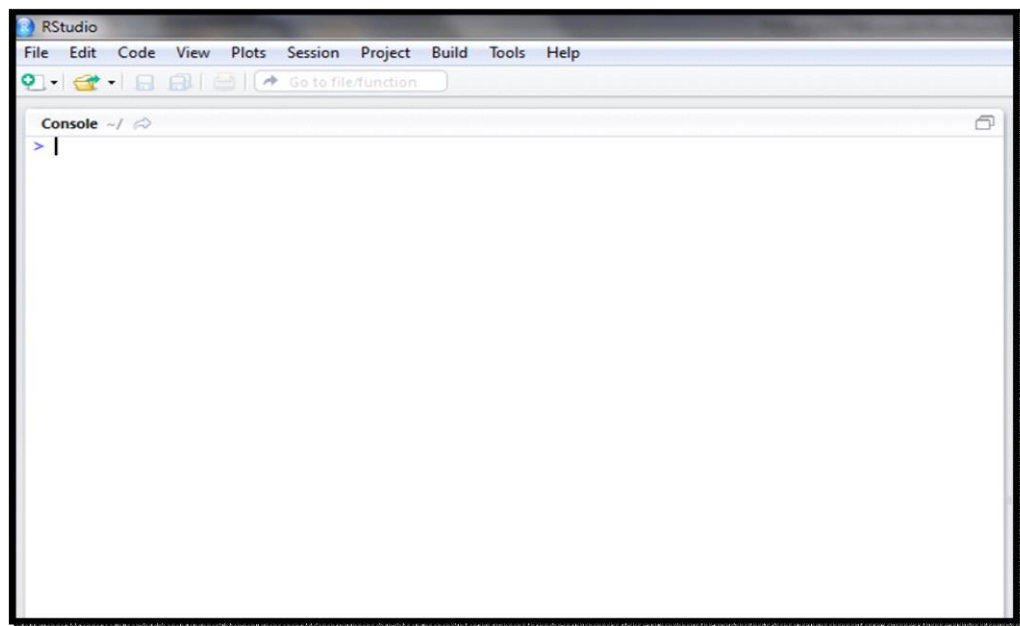


Figura 1 - Tela do RStudio mostrando janela onde comandos podem ser executados.

Um comando no R consiste em uma função (instrução sobre o que fazer) e argumentos que especificam a que a função será aplicada e, em alguns casos, como será aplicada. Os argumentos devem vir entre parêntese e, se houver mais de um, devem ser separados por vírgulas. Por exemplo, a Figura 2 mostra como, por meio do comando “`tolower(x)`”, a função “`tolower`” diz ao R para converter para letras minúsculas todas as palavras contidas em seu argumento `x`, que representa uma estrutura de dados pré-existente.

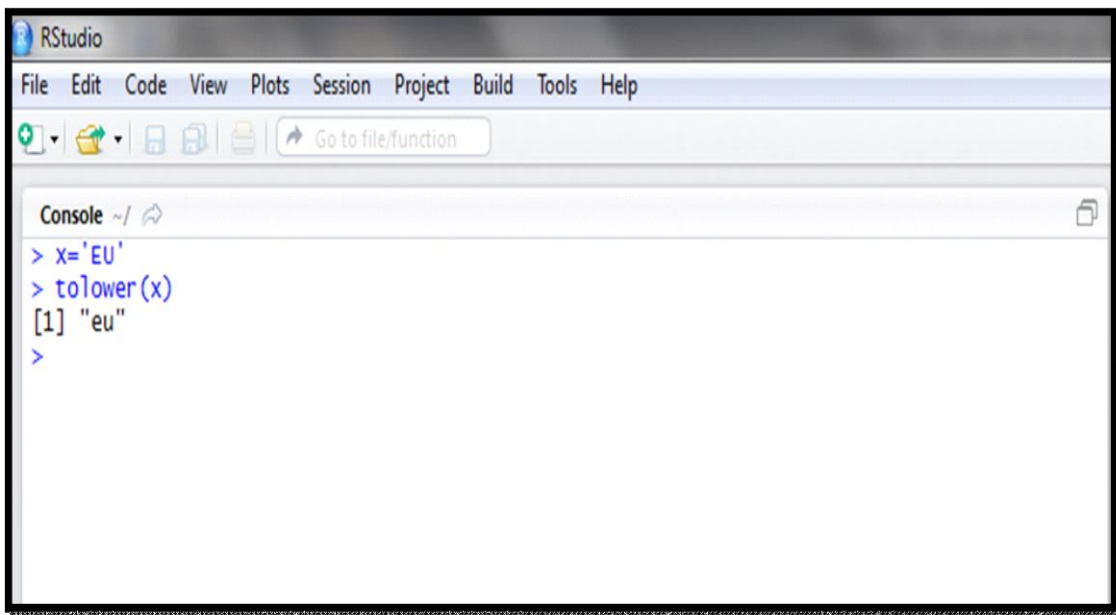


Figura 2 - Resultado da execução do comando "tolower(x)" no RStudio.

Além das funções do R que fazem parte de sua configuração padrão, conjuntos de funções, chamados de “pacotes”, podem ser adicionados à “biblioteca” (library) do R de acordo com os interesses de cada pesquisador. Informações acerca das funções de um pacote são obtidas através da função `library(help=" ")`, colocando-se o nome do pacote entre as aspas. Para instalar e baixar um pacote, utiliza-se as funções `install.packages()` e `library()`, respectivamente. Os argumentos dessas funções são os pacotes que se deseja. Com os pacotes já instalados, a função `help()` pode ser utilizada para obter ajuda sobre uma função específica. A Figura 3, por exemplo, mostra parte do resultado da execução do comando `help(gsub)` para obtermos ajudar sobre função “*gsub*”.

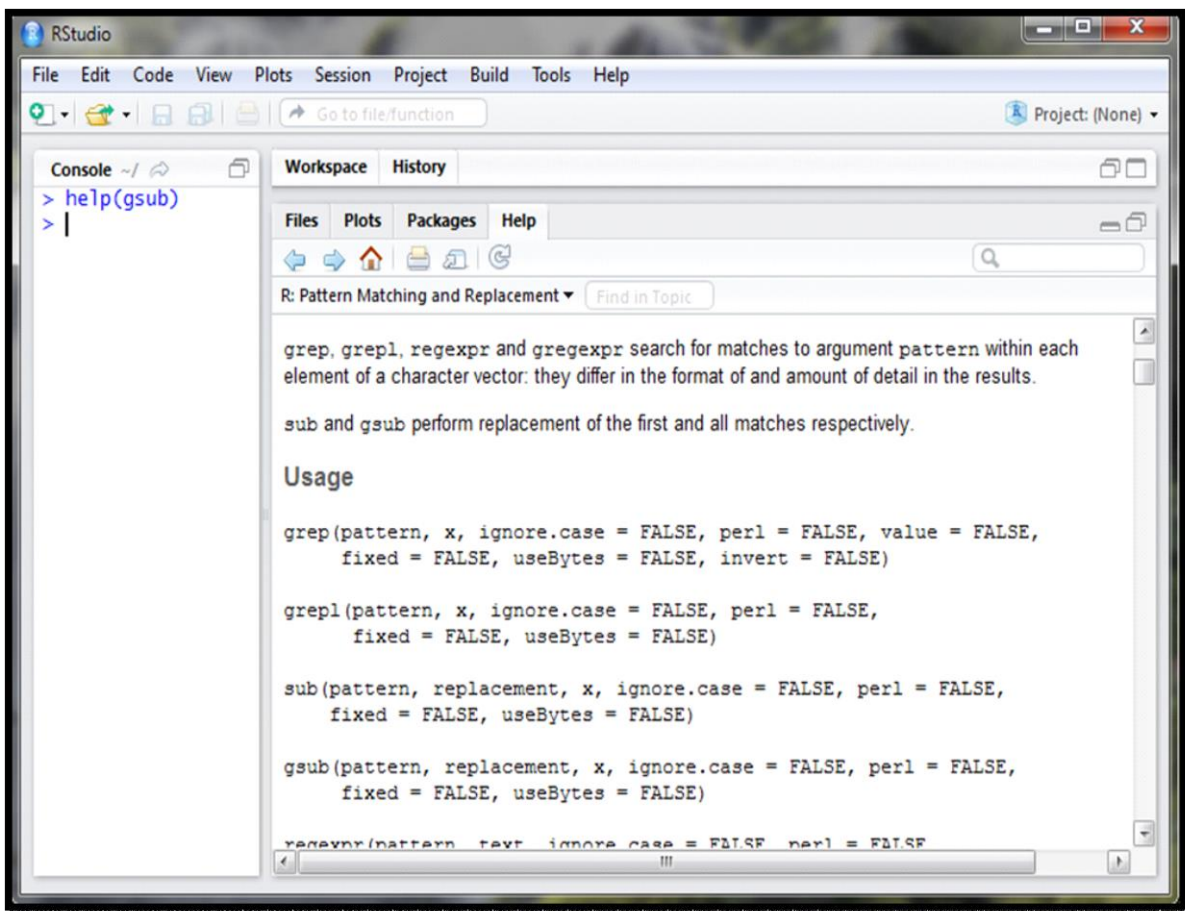


Figura 3 - Resultado da execução do comando "help(gsub)" no RStudio.

Algumas funções presentes nos scripts (conjuntos de comandos) escritos durante este estudo são explicadas abaixo com exemplos de argumentos utilizados:

- `setwd("")`: “setwd” define o diretório de trabalho com o qual se irá trabalhar. Seu argumento especifica o caminho para esse diretório (exemplo: `setwd("C://Users//Júlia//Desktop//Spoken")`).
- `scan(file="", what = , sep = , quiet = )`: “scan” carrega arquivos externos em estruturas de dados do R. Seu primeiro argumento define o arquivo que será carregado. O argumento “what” especifica o tipo de dado a ser lido. Se `what="char"`, o arquivo possui sequências de caracteres. O argumento “sep” define o caractere que irá delimitar a

separação entre os elementos da estrutura de dados (ex: `sep="\n"` significa que o separador é uma nova linha). O argumento `"quiet"` diz se uma linha contendo o número de itens lidos deve ser impressa na janela do R (FALSE) ou não (TRUE).

- `length()`: retorna o tamanho de uma estrutura de dados. Seu argumento é a estrutura de dados.
- `append()`: adiciona elementos a uma estrutura de dados já existente.
- `grep()`: “grep” faz buscas em estruturas de dados. Um exemplo de comando com `grep` é `grep("dtr", whole.corpus, perl=T, value=T)`. O primeiro argumento da função, que deve vir entre aspas, indica o padrão que queremos buscar (ex: “dtr”). O padrão a ser buscado pode ser indicado também por uma “expressão regular”, como `"^.(?={*.?})"`, que representa uma sequência específica de caracteres que se deseja identificar. Utilizamos neste trabalho expressões regulares compatíveis com a linguagem Perl (GRIES, 2009a, p. 79), e, por isso, usamos o argumento `"perl=TRUE"`. Nessa linguagem, alguns caracteres possuem sentidos não literais, como é o caso do ponto final `"."`, que significa “qualquer caractere exceto uma nova linha”, e do sinal de adição `+`, que significa “uma ou mais ocorrência do caractere anterior”, por exemplo. O caractere `\\` é utilizado quando precisamos especificar que um determinado caractere deve ser interpretado literalmente (ex: `\\+`). A estrutura de dados na qual a busca deve ser feita é indicada pelo segundo argumento (ex: `whole.corpus`). O quarto argumento da função “grep”, `"value"`, especifica se os elementos encontrados devem ser retornados (TRUE) nos resultados ou se apenas as suas posições no vetor devem ser informadas (FALSE).
- `gsub()`: faz substituições. O elemento que deve ser substituído é especificado pelo primeiro argumento, e o elemento que irá substituí-lo é definido pelo segundo argumento. O terceiro argumento indica em qual estrutura de dados a substituição deve ocorrer.
- `table()`: gera listas de frequência. Seu argumento é a estrutura de dados contendo as palavras que devem compor a lista de frequência.

Escrevemos scripts do R para manipularmos os arquivos de forma a prepará-los para serem analisados (ex: eliminando informações irrelevantes para esta pesquisa), para contarmos o número de palavras desses arquivos e para localizarmos, marcarmos e extrairmos as expressões

que desejávamos analisar. Uma vez que lidamos com corpora de formatos diferentes (*corpus* etiquetado *versus corpus* não etiquetado; texto disposto horizontalmente *versus* texto disposto na vertical com 1 palavra por linha), diferentes scripts foram necessários para as análises dos diferentes corpora. Como ilustração, explicamos abaixo dois scripts utilizados durante a análise do subcorpus *Written*: um para a contagem do número de palavras e outro para a localização dos verbos que poderiam estar sendo utilizados nas construções de interesse desta pesquisa.

3.2.1 Contagem de palavras do subcorpus *Written* do ICE-GB

Contar o número total de palavras de cada subcorpus é importante para compararmos os diferentes subcorpora considerando suas diferentes extensões. Na Figura 4 pode-se visualizar o script utilizado para a contagem de palavras do subcorpus *Written* do ICE-GB:

```
setwd("C://Users/Júlia/Desktop//Written")
arquivos<-dir("//Users/Júlia/Desktop//Written")
whole.corpus<-vector()
for (i in arquivos) {
  passo.1<-scan(i, what="char", sep="\n", quiet=TRUE)
  passo.2<-grep("\\{", passo.1, perl=TRUE, value=TRUE)
  passo.3<-grep("PUNC", passo.2, perl=TRUE, value=TRUE)
  words<-passo.2[!(passo.2 %in% passo.3)]
  whole.corpus<-append(whole.corpus, words)
  cat(i, length(whole.corpus), "\n", file="word-count.txt", sep="\t", append=TRUE)
}
```

Figura 4 - Script para a contagem de palavras do subcorpus *Written* do ICE-GB.

Através da função “setwd”, definimos o diretório de trabalho com o qual iríamos trabalhar. Depois carregamos todos os arquivos do subcorpus *Written* em uma estrutura de dados no R, para a qual demos o nome de “arquivos”. Em outras palavras, atribuímos o valor “arquivos de C:/Users/Júlia/Desktop/*Written*” à variável “arquivos”. Para a atribuição de valores a variáveis no R, digita-se o símbolo “<-” ou “=” após o nome dado à variável seguido do conteúdo que será armazenado na estrutura de dados. Assim, escrevemos no R:

```
> arquivos<-dir("/Users/Júlia/Desktop/Written")
```

ou

```
> arquivos=dir("/Users/Júlia/Desktop/Written")
```

Em seguida, criamos uma estrutura de dados que foi reservada para conter o resultado da união de todos os arquivos do *subcorpus*:

```
> whole.corpus<-vector().
```

Para dizermos ao R para executar uma série de funções em cada arquivo contido em “arquivos”, usamos o que é chamado de “loop”, indicado pela função “for”:

```
> for (i in arquivos) {
```

O R interpreta a linha acima como: “para cada elemento (i) da estrutura de dados *arquivos* execute os comandos a seguir”. O primeiro comando que colocamos após “{” indica que o R deve carregar (função *scan*) cada elemento de “arquivos” em uma estrutura de dados para qual demos o nome de “passo.1”:

```
> for (i in arquivos) {  
  passo.1<-scan(i, what="char", sep="\n", quiet=TRUE)
```

Selecionamos depois apenas as palavras dentre os elementos contidos nos arquivos de *corpus*. Uma vez que as palavras do *corpus* ICE-GB são colocadas entre {} e ocorrem uma por linha (Fig. 5), a solução utilizada foi localizar e extrair apenas as linhas que continham o caractere “{”. Para isso, utilizamos a função de buscas “grep” e armazenamos os resultados da busca na estrutura de dados “passo.2”:

```

P,PREP(ge) {in}
PC,NP
DT,DTP
DTCE,ART(def) {the}
NPHD,N(com,sing) {past}
AJPO,PP
P,PREP(ge) {than}
PC,AVP(ge)
AVHD,ADV(ge) {at present}
PUNC,PUNC(per) {.}

```

Figura 5 - Parte de um arquivo do *corpus* ICE-GB.

```
> passo.2<-grep("\\\\{", passo.1, perl=TRUE, value=TRUE)
```

Podemos ver na Figura 5 que os sinais de pontuação também aparecem entre {.}. Assim, foi necessário eliminar do resultado armazenado em “passo.2” os sinais de pontuação. Para isso, primeiramente criamos um vetor com os sinais de pontuação (passo.3), e, em seguida, criamos um novo vetor (words) com os elementos de “passo.2” menos os elementos do vetor com os sinais de pontuação:

```

> passo.3<-grep("PUNC", passo.2, perl=TRUE, value=TRUE)
> words<-passo.2[!(passo.2 %in% passo.3)]

```

Finalmente, o resultado de todas as operações supracitadas em cada arquivo do *subcorpus* foi salvo através da função “append” em uma só estrutura de dados, cujo tamanho foi calculado pela função “length”. A função “cat” imprimiu o resultado de “length” no arquivo “word-count.txt”, indicando o número de palavras do *subcorpus*.

```

> whole.corpus<-append(whole.corpus, words)

> cat(i, length(whole.corpus), "\n", file="word-count.txt", sep="\t", append=TRUE)

```

Após definirmos os tamanhos das amostras dos corpora selecionados, cuidamos da

extração das construções que queríamos analisar, como passamos a explicar a seguir.

3.2.2 Como o R facilitou a extração das construções dativas do subcorpus *Written*

Para extrairmos as expressões do subcorpus *Written* a fim de respondermos nossas perguntas de pesquisa, mais de um passo foi necessário. Primeiramente, utilizando o R, localizamos os verbos que provavelmente estavam interagindo com a construção ditransitiva, marcados no ICE-GB com a etiqueta <ditr> (Figura 6). Em seguida, geramos no R uma lista de frequência desses verbos.

```
MVB,V(ditr,edp) {given}
MVB,V(ditr,edp) {told}
MVB,V(ditr,edp) {asked}
MVB,V(ditr,infin) {give}
MVB,V(ditr,infin) {give}
  MVB,V(ditr,past) {gave}
  MVB,V(ditr,edp) {given}
MVB,V(ditr,pres) {permits}
MVB,V(ditr,pres) {shows}
  MVB,V(ditr,ingp) {giving}
MVB,V(ditr,infin) {give}
MVB,V(ditr,edp) {dealt}
  MVB,V(ditr,infin) {offer}
```

Figura 6 - Reprodução de parte das informações contidas no arquivo gerado com o resultado da execução de um script com a função “grep” para extrair verbos marcados como “ditransitivos” (ditr) no subcorpus *Written*.

Era necessário analisarmos os contextos de ocorrência dos verbos para decidirmos se as expressões nas quais eles ocorriam seriam selecionadas para mais análises ou não segundo os critérios especificados mais abaixo. Como essa finalidade, preparamos o arquivo com os textos do subcorpus *Written* eliminando informações indesejadas e transformando os textos verticalizados com 1 palavra por linha em textos ininterruptos com palavras na horizontal. O Quadro 13 mostra a formatação dos arquivos antes e depois de sua manipulação através do R.

Quadro 13 - Comparação entre as formatações do arquivo do subcorpus *Written* antes e depois de sua manipulação através de scripts do R

Formatação original do arquivo com textos do subcorpus <i>Written</i> do ICE-GB	Resultado da manipulação do arquivo para análise
PU,CL(main,ditr,pres) A,PP P,PREP(ge) {In} PC,NP NPHD,N(prop,sing) {FR} SU,NP DT,DTP DTCE,ART(def) {the} NPHD,N(com,sing) {professor} VB,VP(ditr,pres,semi) OP,AUX(semi,pres) {fails to} MVB,V(ditr,infin) {give} OI,NP DT,DTP DTCE,PRON(poss,sing) {his} NPHD,N(com,sing) {monster} OD,NP DT,DTP DTCE,ART(def) {the} NPHD,N(com,sing) {benefit} NPPO,PP P,PREP(ge) {of} PC,NP DT,DTP DTCE,ART(def) {the} NPHD,N(com,sing) {doubt}	In FR the professor fails to give his monster the benefit of the doubt

Durante a leitura dos contextos de ocorrência dos verbos identificados na lista de frequência gerada, observamos algumas ocorrências de construções dativas de interesse desta pesquisa com verbos não etiquetados como (ditr). Por exemplo, na Figura 7 mostramos parte de um arquivo do subcorpus onde uma forma do verbo *render* não etiquetada como “ditr” é usada na construção ditransitiva: “*rendered her a suitable image sor male fantasy*”.

DTCE,ART(def)	{the}	AJHD,ADJ(ge)	{innocent}	PUNC,PUNC(comma)	{,}
NPHD,N(com,sing)	{child}	GENF,GENM	{'s}	NPHD,N(com,sing)	{face}
P,PREP(ge)	{of}	NPHD,N(prop,sing)	{Marilyn Monroe}	MVB,V(cxtr,past)	{ rendered }
NPHD,PRON(pers,sing)	{her}	DTCE,ART(indef)	{a}	AJHD,ADJ(ge)	{suitable}
NPHD,N(com,sing)	{image}	P,PREP(ge)	{for}	AJHD,ADJ(ge)	{male}
NPHD,N(com,sing)	{fantasy}	PUNC,PUNC(per)	{.}		

Figura 7 - Parte de um arquivo do subcorpus *Written* com destaque para o verbo “render”.

Para evitar que dados como esse fossem perdidos, extraímos através de outro script do R todos os verbos do subcorpus, selecionamos aqueles compatíveis com a noção de transferência literal ou metafórica, e analisamos o contexto de uso de cada um deles. Após esse procedimento, seis sentenças com verbos não etiquetados como (ditr) que apresentavam o <recipiente> e o <tema> de eventos de transferência foram selecionadas:

18) Sentenças não etiquetadas como (ditr) no *Written* que foram selecionadas:

- a) (...) I deny any value to the sacred imagery of the Patriarchy (...)
- b) The series of events beginning in the late fourth century (...) provide an important background to the question (...)
- c) Frankenstein fails to provide even the slightest education for his monster.
- d) (...) Pierce had it and is giving it to Oedipa (...)
- e) (...) the innocent, child's face of Marilyn Monroe rendered her a suitable image for male fantasy.
- f) (...) diplomas for Breton monasteries, which conveyed to these Breton communities the same legal status and privileges as many Frankish churches enjoyed.

Os exemplos de scripts acima ilustram como o R foi utilizado nesta pesquisa. Mesmo nos casos dos corpora não-etiquetados, o R facilitou a manipulação dos arquivos de corpora, a localização de informações e a realização de cálculos. Por exemplo, na análise do Br-ICLE, o R possibilitou a geração de uma lista de frequência de palavras a partir da qual foram selecionadas

palavras que poderiam ser verbos utilizados nas construções dativas do inglês. Mais scripts do R serão explorados nas próximas seções, quando explicamos outros aspectos das análises feitas.

Ainda sobre a análise dos corpora do inglês e da interlíngua, definimos na seção 3.3 os critérios para a seleção de expressões identificadas nesses corpora através da ajuda do R.

3.3 Critérios para a seleção de construções dos corpora do inglês e da interlíngua

Critérios sintáticos e semânticos guiaram a extração de sentenças dos corpora da interlíngua e do inglês durante a leitura dos contextos de ocorrência dos verbos selecionados como explicado acima. No que tange à estrutura sintática, as sentenças deveriam possuir um verbo seguido por uma expressão interpretada como <recipiente> e outra como <tema> de eventos de transferência de posse, ambas iniciadas por um sintagma nominal, seja na ordem <Sujeito_{agente} Verbo Objeto_{Recipiente} Objeto_{tema}>, na ordem <Sujeito_{agente} Verbo Objeto_{tema} Oblíquo_{Recipiente}> ou na ordem <Sujeito_{agente} Verbo Oblíquo_{Recipiente} Objeto_{tema}>, como em *I wrote Alice a letter*, *I wrote a letter to Alice* e *I wrote to Alice a letter*, respectivamente. Assim, sentenças que não satisfaziam tais condições foram eliminadas, como nos exemplos 19a-f:

19) Exemplos de sentenças eliminadas da pesquisa:

- a) “We question whether Bishop Lowth is telling the truth”. → verbo + tema
- b) “I told you that I was required (...)” → verbo + recipiente + that-clause
- c) “Sue told me about your (...)” → verbo + recipiente + about
- d) “(...) what I was told by several (...)” → construção passiva
- e) “Her husband told her not to attend (...)” → verbo + NP + infinitive
- f) “(...) tell me if it is (...)” → verbo + NP + if-clause

Foram excluídas também sentenças do *corpus* oral de inglês/L1 com recipientes e/ou temas que foram marcados como não compreendidos durante a transcrição do *corpus* (19g):

- g) “I haven't given them <unclear-words> <unclear-words>”

Outro critério para a seleção de sentenças foi de nível semântico: o referente do sintagma preposicional da construção ditransitiva preposicionada deveria ser semanticamente compatível com o papel de <recipiente>, literal ou metafórico, ou de <beneficiário>, em oposição a uma leitura apenas locativa desse constituinte (20a-b):

20) Exemplos de sentenças eliminadas nas quais o sintagma preposicionado é interpretado como <locativo>:

a) “I had the feeling Peter deliberately brought Lynne to **the house** <locativo> (...)”

b) (...) “sent it to **the above address** <locativo> (...)”

Além de atenderem às condições acima, as sentenças deveriam conter verbos que interagem com as duas construções dativas do inglês, já que buscamos identificar os fatores que influenciam a seleção de uma dessas construções quando a outra também apresenta chances de ser usada. Os verbos que consideramos interagirem tanto com a construção ditransitiva quanto com a ditransitiva preposicionada foram aqueles que encontramos em ambas as construções nos subcorpora do ICE-GB e/ou em dicionários online.

As expressões produzidas por falantes nativos do inglês e por aprendizes brasileiros de inglês/L2 selecionadas como base nos critérios acima foram analisadas segundo as categorias definidas na seção 3.4.

3.4 Definição das categorias para a análise das expressões extraídas dos corpora do inglês e da interlíngua

Após a extração dos dados do inglês/L1 e L2, cada sentença selecionada foi analisada segundo categorias sintáticas e semânticas. Do ponto de vista sintático, as expressões foram classificadas como:

- Ditransitivas (ditr.): quando apresentavam a ordem <Sujeito_{agente} Verbo Objeto_{Recipiente} Objeto_{tema}>, como em “(...) *having a university degree in Brazil does not give people the guarantee of a job (...)*” (Br-ICLE).
- Ditransitivas preposicionadas (ditr. prep.): quando apresentavam a ordem <Sujeito_{agente} Verbo Objeto_{tema} Oblíquo_{Recipiente} >, como em “(...) *people work hard to survive and give their money to institutions in order to support orphans (...)*” (Br-ICLE).
- Ditransitivas preposicionadas invertidas (invertida): quando apresentavam a ordem <Sujeito_{agente} Verbo Oblíquo_{Recipiente} Objeto_{tema}>, como em “*We must give for them a second chance (...)*”(Br-ICLE).

No que tange ao aspecto semântico, classificamos primeiramente as sentenças segundo “categorias semânticas gerais das expressões” baseadas nos sentidos dos sintagmas verbais, que poderiam expressar transferência literal ou transferência metafórica, ou que poderiam fazer parte de expressões idiomáticas ou conter verbos leves. Tais categorias foram chamadas por nós de “literal”, “metafórica”, “idiomática” e “leve”. Essa análise foi feita com o intuito de definirmos para cada *corpus* as proporções dos diferentes tipos de “transferência” aos quais nos referimos ao longo desta tese e assim podermos caracterizar os universos analisados nesse sentido.

Achamos pertinente definir o que entendemos por “sentido literal”, “sentido metafórico”, “expressão idiomática” e “verbo leve”. Existem diversos significados atribuídos à palavra “literal” (LAKOFF, 1986, p. 2). Nesta tese, “literal” refere-se ao significado mais básico (geralmente físico) de uma palavra (LAKOFF, 1993). Por exemplo, na sentença “I gave him a birthday card”, o sentido do verbo *give* é literal porque a sentença expressa uma transferência física de um objeto de um agente para um recipiente. Por outro lado, a sentença “I told you a story” representa uma transferência metafórica. Metáforas correspondem a processos cognitivos através dos quais esquemas de um domínio conceitual (domínio fonte) são mapeados para esquemas de outro domínio (domínio alvo) (PALMER, 1996). Assim, quando o verbo *tell* é utilizado na construção ditransitiva, a expressão resultante requer a interpretação de conceitos de um domínio (informação a ser transmitida de uma pessoa para a outra) através de conceitos de outro domínio (objeto a ser transferido de uma pessoa para a outra). Similarmente, expressões como “Give me a kiss” são licenciadas a partir da metáfora “ações direcionadas a outras pessoas

como objetos transferidos a elas”.

Expressões idiomáticas, em seus exemplos prototípicos, possuem significados que não podem ser completamente inferidos a partir dos sentidos literais das palavras que as compõem, como em “kick the bucket” (NUNBERG et al., 1994). Em outros casos, contudo, é possível estabelecer uma relação entre o significado literal das palavras e o sentido da expressão composta por estas, como em “Her words were a slap in the face” (COOPER, 1998, p. 257). Nesse exemplo, é estabelecida uma analogia entre o efeito de um tapa no rosto e o desapontamento causado por algo que alguém disse. Assim, pode-se dizer que a distinção entre sentido metafórico e sentido idiomático é gradiente, e não categórica. De fato, Gibbs (1992) defende que expressões idiomáticas são motivadas por metáforas conceituais. Às vezes, as metáforas subjacentes às expressões idiomáticas seriam mais transparentes, outras vezes não.

Verbos leves são aqueles que “sofrem um processo de esvaziamento lexical”, de forma que o centro semântico do sintagma verbal é a expressão nominal que acompanha o verbo (MATEUS et al., 2003, p. 312). Em outras palavras, verbos leves são aqueles que, em conjunto com o argumento <tema>, possuem o mesmo significado da forma verbal do mesmo argumento. O Quadro 14 ilustra alguns desses casos.

Quadro 14 - Exemplos de construções com verbos leves

Verbos leves	Formas verbais dos objetos dos verbos leves
“(...) gave him a push (...)”	Pushed him
“(...) gave Swindon a mighty scare (...)”	Scared Swindon
“(...) gave the young couple his blessing (...)”	Blesses the young couple

Marcamos, então, cada sentença selecionada como “literal”, “metafórica”, “idiomática” e “leve”, de forma que, com a ajuda do R, fosse possível calcular facilmente o número de sentenças que pertenciam a cada uma dessas categorias semânticas. A partir do script do R mostrado na Figura 8, carregamos os arquivos com marcações relativas às semânticas dos sintagmas verbais, extraímos as sentenças com tais marcações através da função “grep” e contamos quantas sentenças foram extraídas em cada caso com a função “length”:

```

> x<-scan(file=choose.files(), what="char", sep="\n

> y<-grep("literal", x, perl=TRUE, value=TRUE)
length(y)

> y<-grep("metafórica", x, perl=TRUE, value=TRUE) # selecionar linhas com a marcação “metafórica”
length(y) # retornar a quantidade de linhas selecionadas

> y<-grep("idiomática", x, perl=TRUE, value=TRUE)
length(y)

> y<-grep("leve", x, perl=TRUE, value=TRUE)
length(y)

```

Figura 8 - Script do R utilizada durante análise de aspectos semânticos das sentenças.

Após a análise semântica dos eventos expressos nas sentenças, calculamos as frequências de ocorrência de cada verbo com cada construção dativa. Em seguida, consideramos as associações entre as construções dativas e as seguintes propriedades dos argumentos <tema> e <recipiente> como definidas por Bresnan & Hay (2007):

I. Pronominalidade: refere-se à condição de um argumento ser expresso através de um sintagma cujo núcleo é um pronome pessoal (I.a), um pronome demonstrativo (I.b) ou um pronome indefinido (I.c). Os exemplos que seguem foram retirados do subcorpus *Written* do ICE-GB.

a) “*I 've told you the truth.*”

b) “*Why have I ended up quoting you this?*”

c) “*Isn't it about time you started telling everyone the truth?*”

II. Complexidade sintática: corresponde ao número de palavras. No exemplo (II.a) abaixo, extraído do *Written*, considera-se que o argumento <tema> (5 palavras) é mais extenso do que o argumento <recipiente> (1 palavra):

a) “*(...) to tell her the latest news from Oxford (...)*”

III. Animacidade: refere-se à distinção entre entidades animadas e não-animadas. Entidades animadas incluem humanos, seres humanizados (deuses, fantasmas etc.), expressões cujos referentes denotam um grupo de humanos e animais.

a) “(...) *giving a young person* <animado> *the opportunity to watch those in work* <inanimado>”.

IV. Acessibilidade discursiva: corresponde à qualidade de uma informação ser nova (new) ou dada (given) no discurso. Informações consideradas dadas incluem sintagmas cujos co-referentes localizam-se em uma das 10 linhas antecedentes, além de pronomes pessoais de primeira e segunda pessoa, considerados automaticamente dados no discurso (Ibidem, p. 8).

(a) “(...) *is to grant these early texts* <given> *secondary status* <new> (...)”

Até agora falamos da análise dos corpora do inglês e da interlíngua português/inglês. Na seção 3.5 descrevemos a análise dos corpora do português.

3.5 Análise dos corpora do português brasileiro

Relembramos aqui nossos objetivos em relação à análise dos corpora do português brasileiro: buscar possíveis explicações para a influência do português na interlíngua português/inglês e contribuir com descrições sobre o português do Brasil baseadas em dados de corpora. Para alcançarmos esses objetivos, buscamos a frequência dos padrões do português ilustrados pelas seguintes sentenças extraídas do Humanas:

1. <Sujeito_{agente} Verbo Objeto_{tema} Oblíquo_{Recipiente}>: “(...) eu comecei a dar tinta pra ele (...)”.
2. <Sujeito_{agente} Verbo Oblíquo_{Recipiente} Objeto_{tema}>: “O cara ofereceu pra mim um Vectra completo”.
3. <Sujeito_{agente} Verbo Objeto_{Recipiente} Objeto_{tema}>: “(...) dei ele o alpiste (...)”, “os pesquisadores

deram-lhes camisetas”.

4. < Pronome_{Recipiente} Sujeito_{agente} Verbo Objeto_{tema}>: “(...) me dá mais um pão de queijo aí.”

5. < Sujeito_{agente} Verbo Objeto_{tema} Objeto_{Recipiente}>: “(...) Deus deu dinheiro ocês (...)”.

No intuito de compararmos dados dos corpora da interlíngua e dados dos corpora do português, traduzimos para o português os verbos do inglês encontrados em construções dativas no Br-ICLE, maior *corpus* da interlíngua por nós analisado. O Quadro 15 mostra as traduções feitas com base no Longman Dicionário Escolar. Nove dos dezesseis verbos do português presentes nesse quadro fazem parte da lista dos 100 verbos mais frequentes do C-ORAL-BRASIL (RASO & MELLO, 2012, p. 275).

Quadro 15 - Traduções para o português dos verbos do Br-ICLE selecionados para análise. Verbos marcados com * fazem parte da lista dos 100 verbos mais frequentes do C-oral-Brasil.

Br-ICLE	C-ORAL-BRASIL
Give	Dar*
Bring	Trazer*
Offer	Oferecer
Show	Mostrar*
Teach	Ensinar
Tell	Dizer, Contar*
Cause	Causar
Buy	Comprar*
Hand	Passar*, Entregar
Send	Mandar*, Enviar
Ask	Perguntar*
Build	Construir
Take	Levar*

Devido às diferentes estruturas dos corpora escrito e oral do português, foram utilizados procedimentos diferentes para a localização dos verbos em cada um deles. No caso do C-ORAL-BRASIL, um *corpus* do tipo “anotado”, com apenas um script do R foi possível extrair, de uma só vez, todas as formas de um dado lema. Como mostra o script da Figura 9, o primeiro passo foi

colocar os textos dos 139 arquivos do *corpus* em 1 só arquivo. Em seguida, no “passo 2”, eliminamos as informações indesejadas contidas no *corpus*, manipulamos o conteúdo do arquivo para que ficasse um enunciado por linha, e geramos arquivos que continham apenas enunciados com as formas verbais que estávamos procurando (ex: dar). Repetimos o “passo 2” para cada lema, alterando apenas o primeiro argumento da função “grep”, parte do nono comando desse passo.

```
#Passo 1: para colocar os textos do 139 arquivos em 1 único arquivo
setwd("C:/Users/Julia/Documents/Desktop/coral etiquetado")
arquivos<-dir("C:/Users/Julia/Documents/Desktop/coral etiquetado")
whole.corpus.coral<-vector()#criar vetor para armazenar o resultado
for (i in arquivos) {
  corpus.1<-scan(i, what="char", sep="\n", quiet=TRUE, encoding="UTF-8")
  corpus.2<-paste(corpus.1, collapse="\n")
  corpus.3<-tolower(corpus.2)
  corpus.4<-append(whole.corpus.coral, corpus.3)
  cat(corpus.4, file="whole.corpus.coral.log", sep="\n", append=TRUE)
}

# Passo 2: para colocar cada enunciado em uma linha, eliminar as informações indesejadas e
selecionar as linhas que contenham o lema procurado (exemplo: [dar]).
corpus.pos.1<-scan(file.choose(), what="char", sep="\n")
corpus.pos.2<-gsub("\$;", "[fim]", corpus.pos.1, perl=TRUE)
corpus.pos.3<-grep("\[w+\]", corpus.pos.2, perl=TRUE, value=TRUE)
corpus.pos.4<-gsub("[0-9].*", "", corpus.pos.3, perl=T)
corpus.pos.5<-unlist(strsplit(corpus.pos.4, "^ ", perl=T))
corpus.palavras<-gsub("\[.]*$", "", corpus.pos.5, perl=TRUE)
corpus.palavras.2<-paste(corpus.palavras, collapse=" ")
corpus.palavras.3<-unlist(strsplit(corpus.palavras.2, "[fim]", fixed=TRUE))
dar<-grep("\[dar\b]", corpus.palavras.3, perl=T, value=T)
dar.2<-gsub("\[w+\b]", "", dar, perl=T)
cat(dar.2, file="dar.coral.txt", what="char", sep="\n")
```

Figura 9 - Script do R utilizado para extrair formas de lemas do C-ORAL-BRASIL.

Como ilustração, a Figura 10 mostra as cinco primeiras linhas do arquivo gerado a partir da execução do script acima no R com ocorrências do lema [dar] no C-ORAL-BRASIL. Observe-se que localizamos, de uma só vez, diferentes formas desse lema, como “dá”, “deu” e “dar”.

mas é fechado nu dá pra ir agora

que deu pau que eu acho que deu muito pau nessa taça
a mãe da fafica vai dar o fogão
o tio do anderson vai dar geladeira
a do lado vai dar a televisão

Figura 10 - Parte do resultado do script do R utilizado para a localização de formas do lema [dar] no C-ORAL-BRASIL.

Diferentemente do C-ORAL-BRASIL, o *subcorpus* “Humanas” não é etiquetado. Não foi possível, então, localizarmos todas as ocorrências de um determinado lema com a mesma facilidade. Antes, assim como fizemos com o Br-ICLE, foi necessário gerar uma lista de frequência de palavras, e, a partir dela, buscar ocorrências de formas dos lemas do Quadro 15. Tendo identificado quais formas dos lemas ocorriam no *subcorpus*, buscamos seus contextos de ocorrências no arquivo contendo todos os textos do *subcorpus*.

De forma semelhante ao que fizemos nas análises dos corpora do inglês e da interlíngua, durante a análise dos contextos de ocorrência dos verbos do C-ORAL-BRASIL e do Humanas eliminamos expressões nas quais:

- (I) os verbos não foram interpretados como expressando transferência física ou metafórica (ex: “nu dá pra ir” - interpretado como “não foi possível ir”; “o primeiro táxi deu seis reais” - interpretado como “o primeiro taxi custou 10 reais”; “dar origem” - interpretado como evento de “surgimento” ou invés de “transferência”);
- (II) a reescrita da expressão com outro padrão sintático seria com outra preposição que não “para” ou “a” (Ex: “cê nu me dá nem um beijinho de oi” → cê nu dá nem um beijinho de oi **em** mim);
- (III) ou o <tema> ou o <recipiente> não é exposto após o verbo (Ex: “convida o natalino que ele vai dar uma garrafa de café” / “me dá aí”);
- (IV) o sintagma preposicionado tem interpretação apenas locativa (ex: “e vou comprar um aparelho e vou trazer para a escola”) ou beneficiário não destinatário (ex: “eu levo o arroz lá p’ cê”, significando “eu levo o arroz ao invés de você ter que leva-lo”);
- (V) os verbos não são seguidos de complementos nominais (Ex: “ensinar esse povo a jogar” / “Marília me contou que teve uma época que era proibido (...”).

As expressões selecionadas foram marcadas segundo a numeração abaixo:

- 1: Expressões com a configuração <Verbo Objeto_{Tema} Oblíquo_{Recipiente}>. Exemplo retirado do C-ORAL-BRASIL: “usava ácido bórico e aliás a amiga dela que usava né e deu o nome pra ela”.
- 2: Expressões com a configuração <Verbo Oblíquo_{Recipiente} Objeto_{Tema}>. Exemplo retirado do *corpus* C-ORAL-BRASIL: “a Cassandra deu pra Amanda uma periquita australiana”.
- 3: Expressões com a configuração <Verbo Objeto_{Recipiente} Objeto_{Tema} >. Exemplo do C-ORAL-BRASIL: “dei ele o alpiste”.
- 4: Expressões com a configuração <Pronome_{Recipiente} Verbo Objeto_{Tema} >. Exemplo do C-ORAL-BRASIL: “me dá mais um pão de queijo aí”. Analisar a frequência da estrutura 4 foi considerado por nós interessante para compararmos nossos achados sobre o português/L1 com os resultados de Zara (2009) sobre a transferência dessa configuração sintática para a interlíngua português/inglês.
- 5: Expressões com a configuração < Verbo Objeto_{Tema} Objeto_{Recipiente}>. Exemplo do C-ORAL-BRASIL: “Deus deu dinheiro ocês”.

As duas linhas da Figura 11 exemplificam essa marcação de expressões com números. A primeira delas, marcada com o número 1, foi selecionada para análise através de um script do R com a função “grep”, já a segunda linha, sem nenhuma marcação, não foi. Por meio da função “length”, foi possível contar o número de linhas extraídas.

1 usava ácido bórico e aliás a amiga dela que usava né não e deu o nome pra ela só que nu deu não né

Figura 11 - Exemplificação da marcação feita na análise das sentenças do português.

Relembramos que identificar as frequências das estruturas 1, 2 e 3 acima foi importante para a nossa análise da influência do português nos dados da interlíngua português/inglês

explorados nesta tese. Além disso, dados de frequência de todas as estruturas acima contribuem com estudos anteriores sobre a influência do português na interlíngua português/inglês (ZARA, 2009) e com descrições anteriores do português (SCHER, 1996; GOMES, 2003; BISPO, 2004; BISPO & SALLES, 2005; LUCCHESI & MELLO, 2009). Outros dados de frequência incluídos em nossa análise que contribuem com descrições sobre o português dizem respeito às preposições “a” e “para” nos padrões <Suj. Verb. Obj. Obl.> e <Suj. Verb. Obl. Obj.>.

No que tange à semântica das expressões, verificamos, assim como fizemos na análise dos corpora da interlíngua e do inglês, se as sentenças denotavam transferência literal ou metafórica, se continham expressões idiomáticas ou verbos leves. Exemplos com o verbo *dar* retirados do subcorpus Humanas são fornecidos abaixo para ilustrar a classificação semântica das expressões com base nos sintagmas verbais:

Transferência literal: “(...) as outras deram a espada, o apito e o terço para ela”.

Transferência metafórica: “(...) que vão dando sentido à vida”.

Verbo leve: “(...) também dei destaque a ideias gerais”.

Expressão idiomática: “(...) dando trabalho para seus pais.”

Explicamos acima quais informações buscamos e extraímos dos corpora do português e como as analisamos do ponto de vista sintático e semântico. Na próxima seção tratamos da análise quantitativa dos dados localizados em todos os corpora selecionados para esta pesquisa.

3.6 Análises quantitativas

3.6.1 Cálculo de frequência

A primeira análise quantitativa que fizemos foi em relação às frequências de ocorrência dos padrões sintáticos e de associações entre tais padrões e os fatores explicitados na seção 1.1 desta tese. “Frequência” é o dado quantitativo mais direto que um *corpus* pode fornecer (MCENERY et al., 2006). A frequência “bruta” (raw frequency) corresponde à soma do número

de elementos linguísticos (tokens) que pertencem a uma classificação (type) em um *corpus*. Quando se deseja comparar dois ou mais corpora é necessário, porém, expressar as frequências brutas em termos de uma base comum (ex: 1,000 palavras) para que elas possam ser comparadas (frequência normalizada).

Evison (2010, p. 126) ilustra a importância de se calcular a frequência normalizada a partir da comparação da frequência do pronome “we” em dois subcorpora do BNC: um com 148,624 palavras e outro com 483,913 palavras. No primeiro subcorpus, o pronome ocorre 2,142 vezes, e no segundo, 2,666. Uma vez que os subcorpora possuem tamanhos diferentes, uma comparação direta entre as frequências brutas do pronome “we” em cada subcorpus certamente nos levaria à conclusão equivocada de que ele é um pouco mais frequente no segundo subcorpus citado por Evison. Para que isso não ocorra, podemos, por exemplo, normalizar as frequências em termos de número de ocorrência por 1,000 palavras. Basta dividirmos a frequência bruta pelo número total de palavras do *corpus* e multiplicarmos o resultado por 1,000. No caso do exemplo citado por Evison, os cálculos necessários são:

Subcorpus 1: $(2,142 / 148,624) \times 1,000 = 14,41$

Subcorpus 2: $(2,666 / 483,913) \times 1,000 = 5,50$

Ao contrário da nossa impressão inicial, agora podemos concluir que o pronome “we” é mais que duas vezes mais frequente no primeiro subcorpus do que no segundo.

Damos outro exemplo de como a frequência bruta pode nos induzir ao erro através dos seguintes dados obtidos nesta pesquisa sobre os números de ocorrência de do padrão <Pronome_{Recipiente} Verbo Objeto_{Tema}> no Humanas e no C-ORAL-BRAISL: 123 e 74, respectivamente. Essas frequências brutas nos levariam a dizer que a estrutura em questão é mais frequente no Humanas. Porém, no *corpus* escrito do português esse tipo de construção representa 11,92% das expressões analisadas, enquanto que no C-ORAL-BRASIL ela representa 43,02% de todos os dados. Chegamos, então, a uma conclusão oposta à nossa primeira impressão. Assim, todas as frequências brutas obtidas em nossas análises foram convertidas para frequências relativas em termos de porcentagens considerando o número total do universo de expressões analisadas de cada *corpus*.

3.6.2 Testes de significância estatística e análise de *cluster*

Após obtermos todos os dados de frequência, realizamos diferentes tipos de comparações entre eles. Comparamos (I) duas frequências obtidas em um mesmo *corpus* (ex: frequência da construção ditransitiva *versus* frequência da construção ditransitiva preposicionada no *Written*), (II) duas frequências obtidas em dois corpora distintos (ex: frequência da construção ditransitiva no *Written* e frequência da construção ditransitiva no Br-ICLE) e (III) distribuições de frequências de dois corpora distintos (ex: distribuição de recipientes dados por construção no *Written* e no Br-ICLE).

Quando se quer comparar dados de frequência, a simples observação desses dados pode nos levar a conclusões equivocadas. Por exemplo, ao se observar a tabela fornecida por Gries (2013, p. 5) com as distribuições de tempos e aspectos do inglês em um pequeno *corpus* (Tabela 1), tem-se a impressão de que o “*present tense*” ocorre mais com o aspecto imperfeito, e o “*past tense*” com o aspecto perfeito.

Tabela 1 -Distribuição fictícia de tempos e aspectos do inglês. Fonte: Gries (2003, p. 5).

	Imperfective	Perfective	Totals
Present tense	12	6	18
Past tense	7	13	20
Totals	19	19	38

É necessário reconhecermos, porém, que diferenças observadas entre dados de frequência obtidos em uma amostra podem ser fruto do acaso. Por essa razão, utilizamos técnicas estatísticas chamadas de “testes de significância estatística” ou “testes de hipóteses” para verificarmos se hipóteses sobre as diferenças observadas entre dados de frequência são provavelmente verdadeiras ou provavelmente falsas. Por exemplo, no caso da Tabela 1, para a comparação entre as frequências do *present tense* com o aspecto perfeito e com o aspecto imperfeito, podemos testar as seguintes hipóteses:

- Hipótese nula: a diferença entre 6 e 12 não é estatisticamente significativa.
- Hipótese alternativa: a diferença observada entre 6 e 12 é estatisticamente significativa.

Gries (Ibidem) testou as hipóteses acima por meio do teste de significância estatística chamado de “qui-quadrado”. Com base no resultado desse teste, o autor aceitou a hipótese nula. Portanto, ao invés de afirmarmos que o “*presente tense*” ocorre mais com o aspecto imperfeito, e que o “*past tense*” ocorre mais com o aspecto perfeito, como fizemos inicialmente, teremos que reconhecer que não há diferença estatisticamente significativa entre as frequências 6 e 12 comparadas.

Existem diversos testes de significância estatística que são escolhidos segundo uma série de critérios, como o tamanho da amostra disponível e as frequências mínimas obtidas. Um dos mais difundidos testes de significância estatística utilizados para a análise de dados nominais (dados de frequência de um evento em particular) é o qui-quadrado (Ibidem, p. 165). O teste pode ser executado quando todas as frequências observadas são maiores que 1, quando pelo menos 80% das frequências são maiores que 4 e quando todas as observações são independentes umas das outras (Ibidem, p. 166). Se as condições necessárias para a realização do qui-quadrado não são satisfeitas, o uso do teste Exato de Fisher é recomendado (CORDER & FOREMAN, 2009, p. 179; GRIES, 2013, p. 189; HOWELL, 2011, p. 2), uma vez que esse teste é mais preciso do que o qui-quadrado quando muitas frequências são pequenas (MCDONALD, 2009).

Considerando que os dados obtidos nesta tese variam entre amostras e frequências maiores e menores, optamos pela utilização tanto do teste qui-quadrado quanto do teste exato de Fisher, cada um aplicado aos conjuntos de dados de tese para os quais a literatura os recomenda. Por exemplo, como mostra a Figura 12, para a análise do objeto “tabela” criado no R com as frequências das categorias semânticas de cada *corpus*, utilizamos o qui-quadrado (método=”chisq”) para a análise do *Humanas* (coluna 1), do *Written* (coluna 3) e do *Spoken* (coluna 4), e o teste exato de Fisher (método=”Fisher”) para os demais corpora (colunas 2, 5, 6).

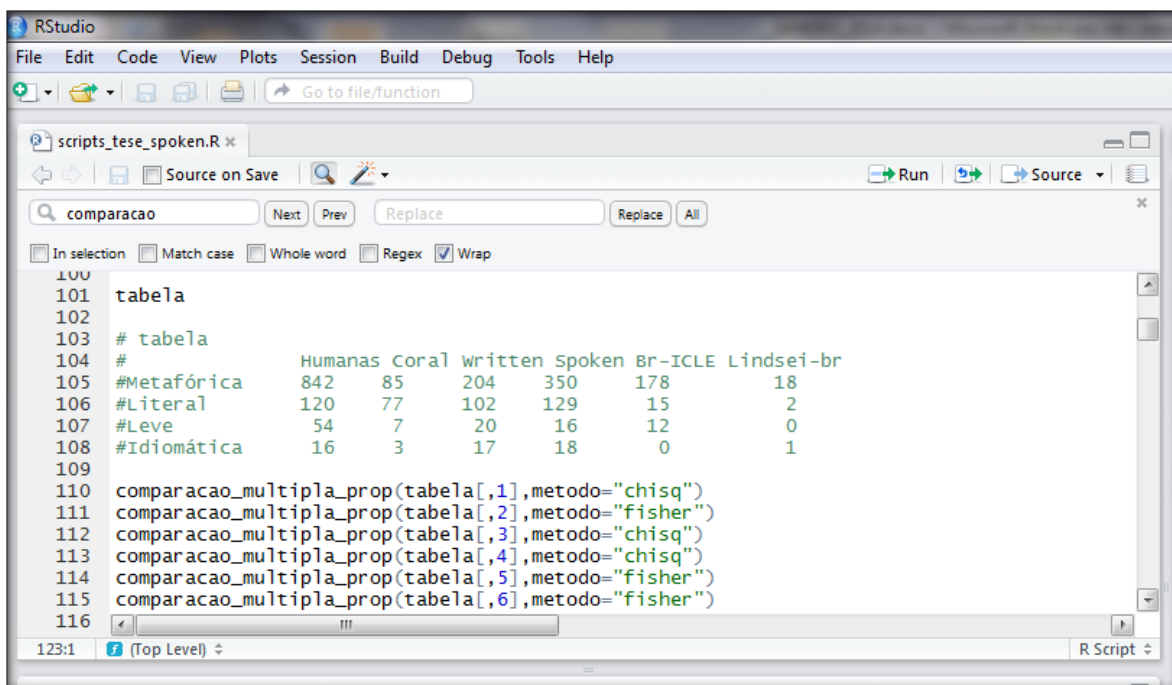


Figura 12 - Exemplos de comandos para a execução dos testes qui-quadrado e exato de Fisher no R.

Em testes de hipóteses como os citados acima, podemos cometer dois tipos de erro. No primeiro deles (erro tipo 1), a hipótese nula é rejeitada quando não deveria ser, e, no segundo (erro tipo 2), a hipótese nula deixa de ser rejeitada quando deveria ser. É necessário, portanto, haver um nível de tolerância para esses tipos de erros. A probabilidade máxima aceitável para o erro tipo 1 é expressa pelo “nível de significância” (α). Não se quer rejeitar a hipótese nula quando ela é verdadeira, então se fixa um nível de significância pequeno para evitar esse tipo de erro, geralmente 5%. Em comparações múltiplas, quando as chances de ocorrência do erro tipo 1 aumentam, recomenda-se ajustar o nível de significância por meio do procedimento de Bonferroni (CORDER & FOREMAN, 2009, p. 81), como fizemos nesta tese. A correção de Bonferroni é realizada com a utilização da fórmula $\alpha_{\beta} = \alpha / K$, sendo α_{β} o nível de significância ajustado, α o nível de significância original e k o número de comparações (Ibidem).

Outra noção importante para compreendermos os testes de significância estatística é a “probabilidade de significância” (p-valor), que corresponde à probabilidade de observarmos um resultado tão extremo ou mais extremo do que de fato foi observado supondo que a hipótese nula seja verdadeira. Um p-valor menor do que o nível de significância indica que seria pouco provável se observar determinado resultado se a hipótese nula fosse verdadeira. Em outras

palavras, se o p-valor for menor do que o nível de significância, a hipótese nula deve ser rejeitada. As etapas da realização de um teste de significância estatística são resumidas abaixo:

- I) As hipóteses a serem avaliadas são formuladas.
- II) O nível de significância é fixado.
- III) O valor-p é calculado e comparado com o nível de significância para a aceitação ou rejeição da hipótese nula.

Os cálculos envolvidos em testes de significância estatística podem parecer complicados para linguistas que não têm familiaridade com estatística, mas basta conhecermos as funções que “dizem ao R como fazer esses cálculos”, e o programa retorna os resultados que precisamos. Por exemplo, a Figura 13 mostra o resultado da execução do comando “`prop.test.exato (c(7,13), c(20,20))`”¹² para a comparação entre as frequências 7 e 13 em um total de 20 ocorrências. Sabendo escrever o comando, precisamos apenas interpretar o que o p-valor obtido nos revela. No exemplo da Figura 13, o p-valor foi igual a 0,1189, portanto a hipótese nula não foi rejeitada.

¹² A função “`prop.test.exato`” foi criada para aplicar o teste qui-quadrado a partir da especificação de duas frequências brutas e da soma dessas frequências (exemplo: `c(7, 13), c(20, 20)`). O nível de significância utilizado é exato.

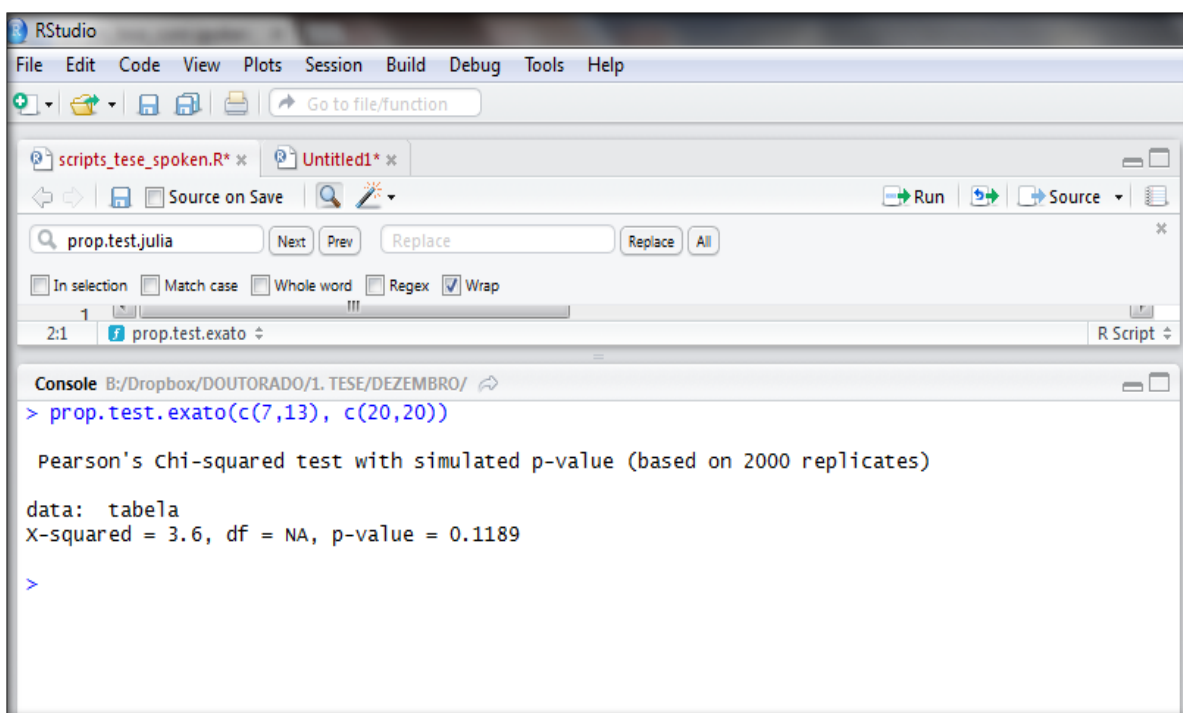


Figura 13 - Resultado da execução de comando com a função `prop.test`, que utiliza o teste qui-quadrado, no R.

O programa R permitiu também que várias análises estatísticas fossem realizadas simultaneamente. Como ilustração, a Figura 14 mostra o resultado da execução de um comando do R com a função “`comparação_multipla.prop`”¹³ utilizado para compararmos, de uma só vez, todas as frequências das categorias semânticas no *corpus Written* entre si. Observe-se que o nível de significância corrigido pela correção de Bonferroni, α_{β} , foi igual a 0.0083.

¹³ Essa função foi criada para esta pesquisa a partir de outras funções existentes no R.

```

RStudio
File Edit Code View Plots Session Build Debug Tools Help
scripts_tese_spoken.R x
Source on Save Run Source
comparacao Next Prev Replace Replace All
Console B:/Dropbox/DOCTORADO/1. TESE/DEZEMBRO/
> tabela
      Humanos Coral Written Spoken Br-ICLE Lindsei-br
Metafórica    842    85    204    350    178    18
Literal        120    77    102    129    15    2
Leve           54     7     20     16    12    0
Idiomática     16     3     17     18     0    1
> comparacao_multiplica_prop(tabela[,3],metodo="chisq")
[1] 8.650853e-15
[1] 3.326598e-50
[1] 3.712742e-52
[1] 6.0778e-16
[1] 2.45684e-17
[1] 0.7353326
$Teste
[1] "Qui-quadrado Exato"

$Resposta
      Literal      Leve      Idiomática
Metafórica "≠ (8.650853e-15)" "≠ (3.326598e-50)" "≠ (3.712742e-52)"
Literal    ""              "≠ (6.0778e-16)" "≠ (2.45684e-17)"
Leve       ""              ""              "≠ (0.7353326)"

$Alpa
[1] 0.008333333
> |

```

Figura 14 - Resultado da execução de um comando do R utilizado para comparar todas as frequências das categorias semânticas no *corpus Written* entre si.

Após utilizarmos os testes de significância estatística para testarmos hipóteses relacionadas a cada propriedade dos argumentos recipientes e temas separadamente conforme explicitado acima, verificamos como os corpora analisados se aproximavam ou se distanciavam um dos outros considerando simultaneamente as proporções de todos os tipos de recipientes e de temas analisados nesta pesquisa (pronome, NP lexical, dado, novo, animado, inanimado, mais extenso, menos extenso, de igual extensão) na construção ditransitiva e na ditransitiva preposicionada. Para esse fim, fizemos uma análise de *cluster*, utilizada para classificar um conjunto de elementos, no nosso caso os corpora, em grupos cujos membros compartilhem propriedades, de forma que os membros de um grupo sejam muito semelhantes entre si e ao mesmo tempo diferentes dos membros de outros grupos (CANTOS GOMEZ, 2002; GRIES,

2013, p. 337).

O formato do resultado de uma análise de cluster é ilustrado na Figura 15. Para explicar o que essa análise pode nos revelar, Gries (Ibidem) lança mão de julgamentos fictícios de falantes nativos do inglês sobre a similaridade entre os seguintes fonemas de sua L1: /b/, /d/, /f/, /g/, /l/, /m/, /n/, /p/, /s/, /t/ e /v/. O autor explica que o diagrama da Figura 15 mostra que seus falantes fictícios dividiram os fonemas em dois grupos maiores segundo a forma de articulação das consoantes: um com consoantes plosivas (/b/, /p/, /t/, /d/, /g/) e outro com consoantes nasais, líquidas e fricativas (/l/, /n/, /m/, /v/, /f/, /s/). Além disso, eles formaram subgrupos que destacam outros tipos de similaridades e diferenças entre as consoantes, como o fato de /b/ e /p/ serem semelhantes entre si, mas diferentes dos outros membros de seu grupo maior por serem bilabiais.

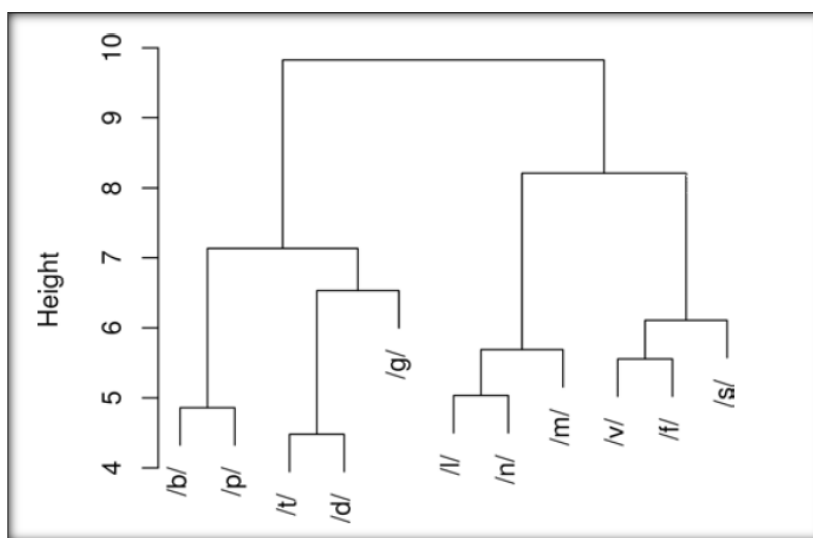


Figura 15 - Resultados fictícios de uma análise de cluster de consoantes do inglês.
Fonte: Gries (2013, p. 337).

Na análise de cluster que fizemos, optamos por usar um *heatmap* juntamente com o dendrograma ilustrado pela Figura 15 acima. O *heatmap* é um recurso gráfico que usa cores para representar dados quantitativos de multi-variáveis que medem diferentes aspectos de um grupo de entidades (FEW, 2006, p. 1). Essa definição de *heatmap* relaciona-se com as características da análise que fizemos como mostramos na Figura 16.

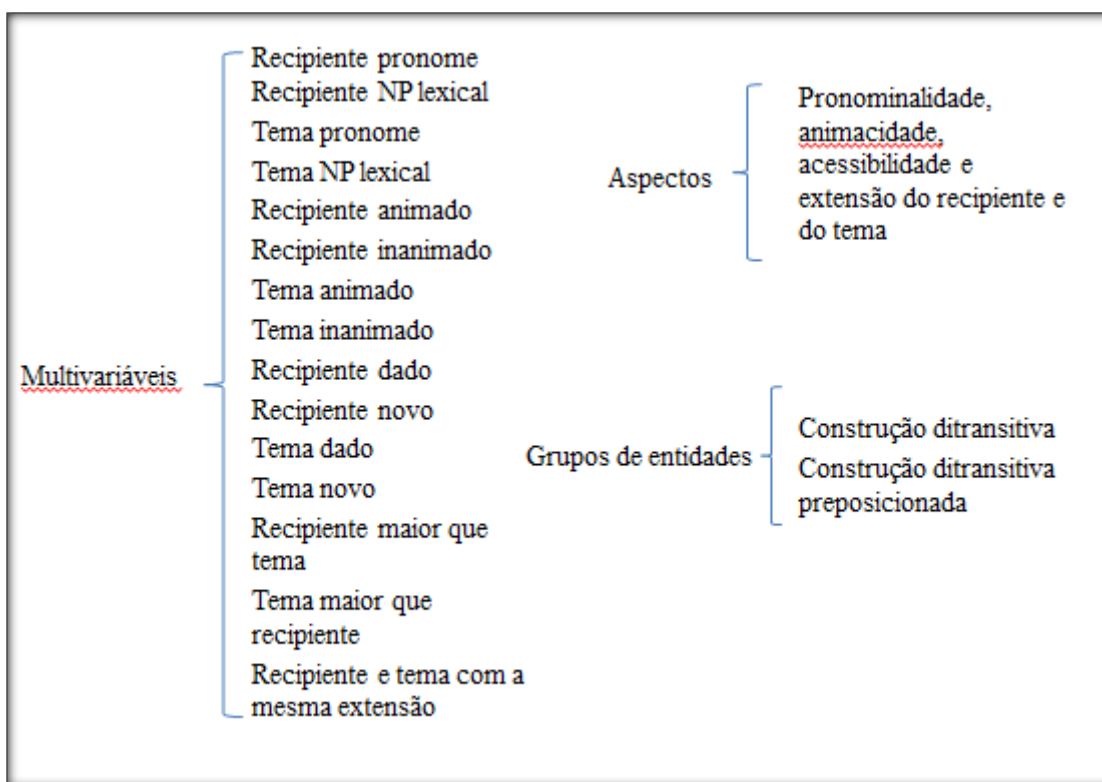


Figura 16 - Relação entre os termos presentes na definição de *heatmap* e as características da análise feita nesta tese.

Um exemplo de *heatmap*¹⁴ é dado na Figura 17. Ele foi utilizado para comparar 17 línguas quanto a 15 aspectos. Esse exemplo evidencia uma das grandes vantagens da análise de cluster: a possibilidade de capturar de forma rápida e confiável semelhanças entre grandes números de dados. No *heatmap* abaixo, cores iguais indicam em quais aspectos as línguas se assemelham.

¹⁴ <http://www.replicatedtypo.com/visualising-language-typology-plotting-wals-with-heat-maps/5189.html>

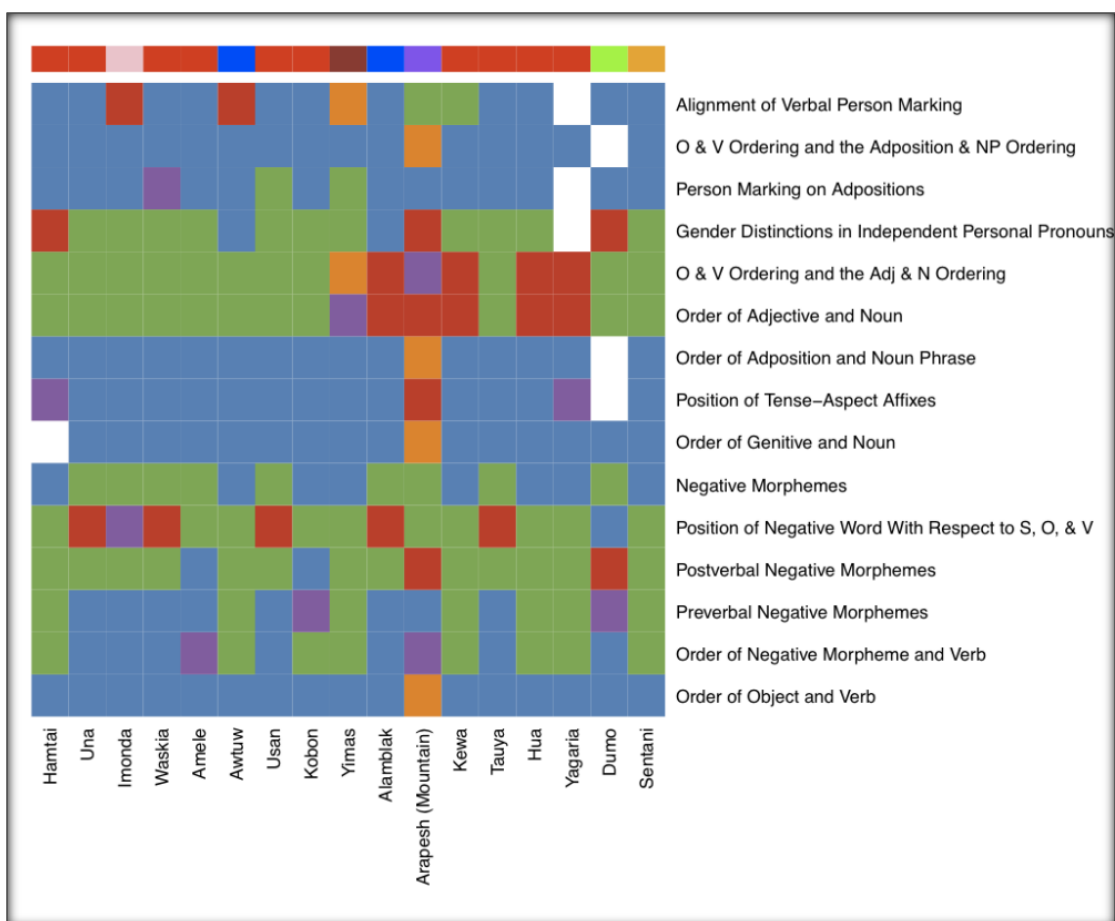


Figura 17 - Exemplo de *heatmap*.
Fonte: LITTAUER (2012).

Neste capítulo detalhamos a metodologia utilizada para alcançar os objetivos desta pesquisa. O próximo capítulo apresenta os resultados por nós obtidos.

4. RESULTADOS E DISCUSSÃO

Neste capítulo apresentamos e discutimos os resultados obtidos por meio da metodologia exposta no capítulo 3. Na seção 4.1, fazemos considerações acerca dos números de expressões selecionadas de cada *corpus* e caracterizamos os universos analisados em termos dos tipos de transferências expressos nas sentenças que deles fazem parte. Na seção 4.2, mostramos os resultados relativos às distribuições dos padrões frasais e preposições nos corpora do português, que representam a L1 dos aprendizes. Na seção 4.3, passamos para os resultados obtidos acerca das distribuições das construções nos corpora de falantes nativos e de aprendizes de inglês. Os resultados relativos à construção ditransitiva preposicionada invertida no Br-ICLE são discutidos separadamente na seção 4.4. Finalizamos este capítulo, na seção 4.5, com o que nossos dados revelam sobre os fatores que influenciam a escolha entre a construção ditransitiva e a ditransitiva preposicionada nos corpora de inglês/L1 e da interlíngua português/inglês.

4.1 Números de expressões selecionadas de cada *corpus* e caracterização dos universos analisados quanto às categorias semânticas gerais das expressões

Os números totais de expressões selecionadas de cada *corpus* conforme detalhamos no capítulo 3 são indicados no Quadro 16. Não surpreendentemente, no geral, quanto maior o *corpus*, maior o número de expressões selecionadas. Como uma exceção a essa regra, o número de expressões selecionadas do C-ORAL-BRASIL foi menor do que o número de expressões selecionadas do Br-ICLE, apesar de aquele ser maior do que este. Se considerarmos uma das diferenças entre os canais oral e escrito no português do Brasil, podemos achar uma explicação para esse resultado. No *corpus* oral, houve uma tendência maior de recipientes e temas de eventos de transferência de posse não serem expressos após o verbo por estarem prontamente acessíveis no contexto extra-linguístico ou no contexto linguístico anterior. Por exemplo, na expressão “o outro vai dar micro-ondas”, retirada do C-ORAL-BRASIL, a pessoa que receberia o micro-ondas

(recipiente) não é mencionada explicitamente, pois já havia sido mencionada anteriormente e faz parte do tópico do diálogo, sendo facilmente recuperada pelos participantes da interação. Vários outros casos como esse foram observados no *corpus* oral. Isso explica a menor ocorrência nesse *corpus* dos padrões focalizados nesta tese, nos quais nem o recipiente nem o tema poderiam estar omitidos após o verbo.

Quadro 16- Número total de expressões selecionadas de cada *corpus* analisado nesta tese

<i>Corpus</i>	Nº de palavras do <i>corpus</i>	Nº de expressões selecionadas
Humanas	1.786.289	1.032
<i>Spoken</i>	637.562	516
<i>Written</i>	423.702	343
C-ORAL-BRASIL	208.130	172
Br-ICLE	160.000	205
LINDSEI-BR	30.952	21

Queremos chamar a atenção para alguns aspectos interessantes que observamos durante a seleção de sentenças dos corpora do português. Primeiramente, como se pode observar na Tabela 2, o número de formas verbais extraídas automaticamente do C-ORAL-BRASIL com um script do *software* R e o número de palavras identificadas como candidatas a formas verbais de interesse desta pesquisa durante a leitura da lista de frequência de palavras do Humanas são muito maiores do que os números de expressões de fato selecionadas para mais análise após a verificação dos contextos de ocorrência dos verbos. Por exemplo, das 858 ocorrências do verbo “dar” no C-ORAL-BRASIL, apenas 97 são nas construções que buscamos nesse *corpus*. Isso mostra que as construções focalizadas nesta tese representam apenas uma pequena amostra das construções que interagem com os verbos que analisamos, evidenciando a variedade de interações possíveis entre construções lexicais e construções sintáticas. Consequentemente, quando se quer analisar certos padrões sintáticos em um *corpus*, é necessário um número grande de dados para que muitas ocorrências desses padrões sejam encontradas.

Tabela 2 - Números de verbos cujos contextos foram analisados e números de expressões selecionadas de cada *corpus* do português

Lema	C-ORAL-BRASIL		Humanas	
	Verbos cujos contextos foram analisados	Expressões selecionadas	Verbos cujos contextos foram analisados	Expressões selecionadas
Dar	858	97 (11,31%)	4.068	595 (14,63%)
Dizer	182	3 (1,65%)	2.780	13 (0,47%)
Passar	415	14 (3,37%)	2.181	16 (0,73%)
Contar	141	14 (9,93%)	1.324	11 (0,83%)
Mostrar	80	5 (6,25%)	1.253	32 (2,55%)
Levar	206	2 (0,97%)	1.000	11 (4,40%)
Trazer	78	5 (6,41%)	596	89 (14,93%)
Causar	3	0 (0,00%)	557	25 (4,49%)
Oferecer	12	7 (58,33%)	438	112 (25,57%)
Construir	23	0 (0,00%)	373	(0,27%)
Perguntar	66	2 (3,03%)	351	1 (0,28%)
Ensinar	38	5 (13,16%)	308	24 (7,79%)
Comprar	197	4 (2,03%)	126	4 (3,17%)
Mandar	121	14 (11,57%)	118	16 (13,56%)
Entregar	41	0 (0,00%)	113	30 (26,55%)
Enviar	0	0 (0,00%)	97	19 (19,59%)
Total	2.461	172 (6,99%)	15.682	1.032 (6,58%)

Chama-nos também a atenção o fato de que, quando calculamos a porcentagem total de verbos selecionados em relação ao número total de verbos inicialmente analisados por *corpus*, encontramos proporções semelhantes entre os dois corpora do português: foram selecionados 6,58% e 6,99% dos verbos analisados no *corpus* escrito e no *corpus* oral, respectivamente (Tabela 2). A aplicação do teste qui-quadrado revelou que essas porcentagens são estatisticamente semelhantes (p-valor = 0,4759). Concluimos, então, que, embora os corpora do português analisados nesta tese tenham tamanhos consideravelmente diferentes, eles são semelhantes quanto à proporção total de verbos que interagem com as 5 estruturas do português buscadas nos corpora em relação ao número total de verbos compatíveis com a noção de transferência em cada *corpus*. Isso indica que um *corpus* do tamanho do C-ORAL-BRASIL pode refletir certos aspectos encontrados em corpora maiores como o Humanas.

Outro resultado da análise dos corpora do português que destacamos diz respeito à

comparação das distribuições de lemas particulares encontradas em cada *corpus* (Tabela 2). A aplicação do teste exato de Fisher revelou que o Humanas e o C-ORAL-BRASIL se diferem um do outro nesse sentido (valor-p = 0,0004), ou seja, que há uma diferença significativa entre os canais escrito e oral quanto à frequência de lemas particulares.

As sentenças selecionadas dos corpora do português e do inglês L1 e L2 foram classificadas em categorias que distinguem os diferentes casos de “transferência”: a transferência física (categoria “física”), a transferência metafórica (categoria “metafórica”), a transferência expressa por um verbo leve (categoria “leve”) e a transferência presente em uma expressão idiomática (categoria “idiomática”). Os dados de frequências brutas exibidos na Figura 18 mostram que, quando falamos de “transferência” nos referindo aos dados que obtivemos, estamos tratando na maioria das vezes da transferência metafórica.

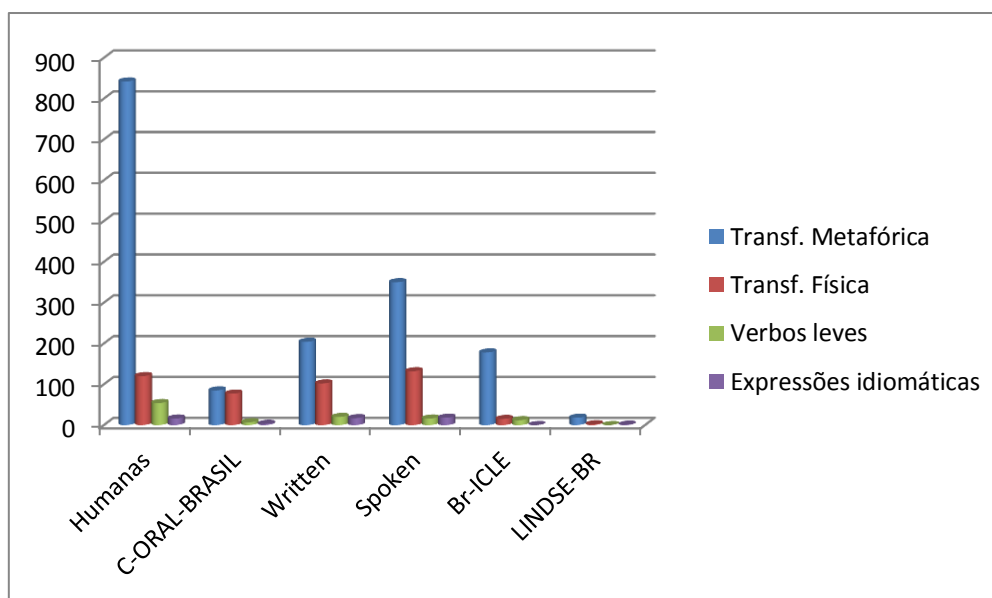


Figura 18 - Distribuição das categorias semânticas gerais das expressões por *corpus*.

A aplicação dos testes qui-quadrado e teste exato de Fisher para comparar a frequência da categoria “metafórica” com as demais categorias semânticas por *corpus* resultou nos p-valores exibidos na Tabela 3. Interessantemente, apenas no *corpus* oral do português encontramos outra

categoria semântica estatisticamente tão frequente quanto a categoria “metafórica”: a categoria “literal” (p-valor = 0,4496, em destaque na tabela). Os tipos de interações presentes no C-ORAL-BRASIL podem ter favorecido a maior menção a eventos de transferência física de posse nesse *corpus*. De fato, muitos são os fatores que podem influenciar dados de frequências observados em corpora.

Tabela 3 – P-valores resultantes da aplicação do teste qui-quadrado (Humanas e *Written*) e do teste exato de Fisher (Br-ICLE e LINDSEI-BR) para a comparação das frequências das categorias semânticas por *corpus*.

Humanas		Literal	Leve	Idiomática
	Metafórica	$\neq 4,2083\text{e-}22$	$\neq 1,2697\text{e-}267$	$\neq 3,3771\text{e-}297$
C-ORAL-BRASIL		Literal	Leve	Idiomática
	Metafórica	$= 0,4496$	$\neq 2,1243\text{e-}23$	$\neq 1,3827\text{e-}27$
<i>Written</i>		Literal	Leve	Idiomática
	Metafórica	$\neq 8,6508\text{e-}15$	$\neq 3,3265\text{e-}50$	$\neq 3,7127\text{e-}521$
<i>Spoken</i>		Literal	Leve	Idiomática
	Metafórica	$\neq 4,0335\text{e-}43$	$\neq 1,9522\text{e-}104$	$\neq 5,7837\text{e-}103$
Br-ICLE		Literal	Leve	Idiomática
	Metafórica	$\neq 9,1716 \text{e-}42$	$\neq 4,1679\text{e-}104$	$\neq 1,2292\text{e-}102$
LINDSEI-BR		Literal	Leve	Idiomática
	Metafórica	$\neq 1,1044\text{e-}06$	$\neq 7,5205\text{e-}09$	$\neq 1,2596\text{e-}07$

$\alpha_{\beta} = 0,0083$

Entender a maior ocorrência geral da categoria “metafórica” nos universos por nós analisados não é difícil. Primeiramente, alguns verbos desta pesquisa necessariamente expressam transferência metafórica, como “*tell*” e “*show*”. Além disso, há muitas possibilidades de usos metafóricos de verbos que podem expressar transferência física, como no caso de “*give*” em “*give*

me the opportunity”.

Após compararmos as frequências das categorias semânticas por *corpus*, comparamos os corpora entre si quanto às distribuições das categorias semânticas por meio da aplicação do teste exato de Fisher (Tabela 4). Encontramos diferença entre os corpora escrito e oral do português (valor-p = 0.004), mas não entre os corpora escrito e oral do inglês (p-valor = 0,0329). É possível, então, que no português, mas não no inglês, o canal influencie a frequência das categorias semânticas analisadas nesta tese. Considerando agora o mesmo canal, mas línguas diferentes, encontramos diferenças entre o português e o inglês tanto na escrita (p-valor = 0,004) quanto na fala (p-valor = 0,004). Interessantemente, o *corpus* escrito da interlíngua foi estatisticamente semelhante ao *corpus* escrito da L1 (p-valor = 0,1154), mas foi diferente do *corpus* escrito da língua-alvo (p-valor = 0,004). Em outras palavras, em relação à distribuição das categorias semânticas gerais das expressões, quando consideramos o canal escrito, o *corpus* da interlíngua aproxima-se mais do *corpus* do português/L1 do que do *corpus* do inglês/L1. O *corpus* oral da interlíngua foi estatisticamente semelhante a todos os outros corpora. O tamanho da amostra analisada do LINDSEI-BR (21 construções) pode não ter sido suficiente para que diferenças entre esse *corpus* e os demais fossem notadas.

Tabela 4 - Resultados da aplicação do teste exato de Fisher para verificar se as distribuições das categorias semânticas são semelhantes entre os corpora. Destaque para p-valores que levam à aceitação da hipótese nula.

	C-ORAL-BRASIL	Written	Spoken	Br-ICLE	LINDSEI-BR
Humanas	≠ 0,0004	≠ 0,0004	≠ 0,0004	= 0,0809	= 0,4362
C-ORAL-BRASIL		= 0,0069	≠ 0,0004	≠ 0,0004	= 0,0074
Written			= 0,0394	≠ 0,0004	= 0,0829
Spoken				≠ 0,0004	= 0,2978
Br-ICLE					= 0,0934

$\alpha_{\beta} = 0,0033$

Definimos nesta seção os tamanhos dos universos analisados. Caracterizamos também os corpora em termos das proporções dos diferentes tipos de transferência expressos nas sentenças que os compõem, enriquecendo, assim, a descrição dos corpora comparados. Passemos agora para os resultados relativos às frequências das construções sintáticas nos corpora, começando pelos corpora do português, no quais analisamos também as frequências das preposições introdutórias do dativo “a” e “para”.

4.2 Resultados dos corpora do português

As distribuições dos padrões de 1 a 5 nos corpora escrito e oral do português analisados são apresentadas na Tabela 5. O gráfico da Figura 19 destaca a comparação entre as frequências brutas dessa tabela.

Tabela 5 - Distribuição das construções nos corpora do português

Construção / Exemplo	Humanas	C-ORAL-BRASIL
1. Ditransitiva preposicionada Maria deu um livro para João.	58,23% (601)	45,93% (79)
2. Ditransitiva Preposicionada invertida Maria deu para João um livro.	22,67% (234)	7,56% (13)
3. Ditransitiva Maria deu João um livro.	7,18% (74)	2,91% (5)
4. Pronome Verbo Maria lhe deu um livro.	11,92% (123)	43,02% (74)
5. Maria deu um livro João.	0,00% (0)	0,58% (1)

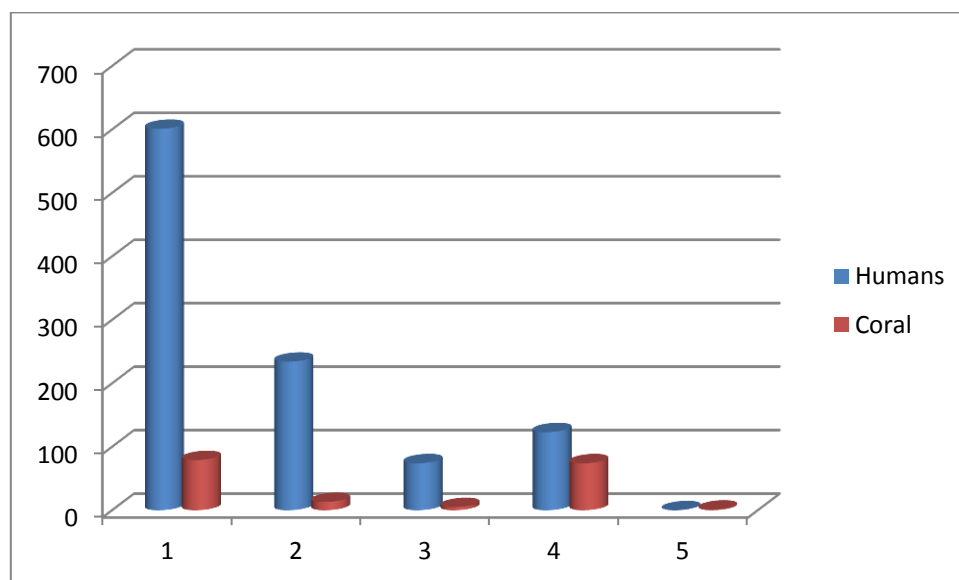


Figura 19 - Gráfico com as distribuições dos padrões de 1 a 5 nos corpora escrito e oral do português analisados.

Os dados apresentados acima muito nos mostram. Primeiramente, comparemos as frequências das construções 1 e 2 para adicionarmos informações ao que Gomes (2003) já relatou sobre essas construções. Em sua pesquisa, a autora constatou que a construção ditransitiva preposicionada (estrutura 1) representou a configuração mais frequente para complementos de verbos bitransitivos. Contribuímos agora com a informação de que a construção ditransitiva preposicionada é não apenas mais frequente, mas sim *significativamente* mais frequente do que a ditransitiva preposicionada invertida (estrutura 2) tanto no *corpus* escrito ($p\text{-valor} = 1,5164e-60$) quanto no *corpus* oral ($p\text{-valor} = 2,2603e-15$). Na verdade, podemos dizer que a ditransitiva preposicionada invertida é infrequente na fala (7,56%), sendo mais associada à escrita ($p\text{-valor} = 0,0004$). Para usarmos a terminologia de Chater e Manning (2006), podemos dizer que, na fala informal do PB, a construção ditransitiva preposicionada invertida é linguisticamente possível, mas não provável.

Relacionando os resultados exibidos na Figura 19 com dados da interlíngua português/inglês que mostram que aprendizes brasileiros do nível de proficiência iniciante em inglês/L2 já demonstram conhecimento sobre a construção ditransitiva preposicionada do inglês (ZARA, 2009), podemos levantar a hipótese de que construções da estrutura argumental da L2

que possuem correspondentes frequentes na L1 são adquiridas mais rapidamente pelos aprendizes. Ainda estabelecendo um paralelo entre dados da L1 com dados da interlíngua, a baixa frequência da construção 2 (preposicionada invertida) observada tanto no *corpus* escrito quanto no *corpus* oral do português analisados nesta tese pode ser um dos fatores que explicam porque Zara (2009) não identificou uma tendência para a transferência desse padrão do português/L1 para a interlíngua português/inglês em uma tarefa de julgamento de aceitabilidade de sentenças.

Focalizemos agora os dados relativos à estrutura 3 (construção ditransitiva). Observamos que tanto na variedade do português falado informal, representada pelo C-ORAL-BRASIL, quanto na escrita “cultura” do português, representada pelo Humanas, a ocorrência da construção ditransitiva foi muito baixa, representando 2,91% e 7,18% das expressões analisadas em cada *corpus* respectivamente. Vale ressaltar que, no *corpus* escrito, todas as ocorrências da ordem <Sujeito Verbo Objeto_{recipiente} Objeto_{tema}> foram com recipientes expressos através de pronomes oblíquos átonos, como em “Eu próprio enviei-lhe um fax, a contragosto, pressionado pelo comitê de apoio (...)”. Logo, em descrições do português sobre formas dativas, deve-se ressaltar que a estrutura <Sujeito Verbo Objeto_{recipiente} Objeto_{tema}> é rara em relação a outros padrões frasais, pelo menos na escrita “cultura” e na fala informal das variedades do português consideradas nesta pesquisa.

Embora a probabilidade de ocorrência da construção ditransitiva em outras variedades do português possa ser maior do que a encontrada em nossa análise de *corpus*, resultados de estudos anteriores também apontam para a baixa ocorrência dessa construção no PB. Por exemplo, no estudo de Lucchesi e Mello (2009) citado na seção 2.6 desta tese, foi encontrada uma frequência da ditransitiva igual a 17%, que, embora seja maior do que a frequência encontrada na nossa pesquisa, não deixa de ser baixa¹⁵. Relacionando esses resultados com nosso interesse sobre a influência da L1 na interlíngua, sugerimos que a baixa frequência da construção ditransitiva no português explique porque brasileiros aprendizes de inglês levam mais tempo para adquirirem a construção ditransitiva do inglês em comparação à ditransitiva preposicionada, como mostra Zara (2009).

Passando agora para estrutura 4, encontramos uma diferença notável entre os canais oral e escrito: esse padrão é estatisticamente mais frequente C-ORAL-BRASIL do que Humanas (p-

¹⁵ Talvez uma frequência da ditransitiva no PB seja encontrada em variedades muito específicas utilizadas por grupos de falantes que compartilham as características apontadas como favorecendo o uso dessa construção, como idade mais avançada e nível de educação formal mais baixa (LUCCHESI & MELLO, 2009).

valor =0,0004). Portanto, nossos resultados indicam que há no PB uma associação maior entre o padrão <Suj. Pronome_{recipiente} Verbo Obj.> e o canal oral do que entre esse padrão e o canal escrito. Esse resultado reflete duas tendências do português: o maior uso de pronomes pessoais na fala do que na escrita¹⁶ e a preferência pelo uso dos pronomes oblíquos em posição pré-verbal (ROCHA, 1998). Assim, temos mais um exemplo de como os canais da linguagem influenciam o uso de estruturas linguísticas.

Olhando ainda para a frequência da estrutura 4, novamente consideramos relevante citar Zara (2009) para fazermos mais uma observação sobre efeitos de frequência no processo de transferência linguística. Vimos acima que a estrutura 4 é tão frequente quanto a 1 no *corpus* oral do português. Devido a essa alta frequência, que um falante nativo do português poderia já supor, Zara (2009) levantou a hipótese de que brasileiros aprendizes iniciantes de inglês aceitariam sentenças como “*He me gave a kiss*” por um processo de analogia com sentenças do tipo “Ele me deu um beijo”. Ao contrário das expectativas da autora, porém, os sujeitos de sua pesquisa tenderam fortemente a rejeitar as sentenças em inglês com a ordem <Suj. Pronome_{recipiente} Verbo Objeto> mesmos nos estágios iniciais de aquisição da L2. O fato de aprendizes tenderem a não transferir para a interlíngua padrões sintáticos da sua L1 não testemunhados na L2 mostram que aprendizes são sensíveis não somente à frequência de estruturas de sua L1, mas também à frequência de estruturas da L2.

Por fim, vejamos os resultados para a estrutura 5. A omissão da preposição apontada por Gomes (2003) em frases como “Eu dei o pão eles” não foi identificada no *corpus* escrito do português e chegou perto de zero (0,58%) no *corpus* oral. Portanto, há uma indicação de que, nos corpora do PB analisados, tal estrutura foi resultante de aspectos não-linguísticos que influenciam a performance de falantes, tais como cansaço, distrações e perda de atenção. Sugerimos ainda, com base em nossa experiência enquanto falantes nativos do português, que a estrutura 5 seja uma característica específica de interações familiares como em “Dá isso mamãe¹⁷”, e, por isso, não ocorra em um *corpus* escrito do português culto como o Humanas. Mais dados empíricos são necessários para confirmar nossas hipóteses sobre o uso da estrutura 5. De qualquer forma,

¹⁶ Ao fazermos uma busca por pronomes pessoais no *Corpus* do Português disponível em <http://www.corpusdoportugues.org/x.asp>, encontramos 42,677.08 ocorrências por milhão na seção “ORAL”, em contraste com 12,484.48 ocorrências por milhão na escrita acadêmica e 16,226.74 ocorrências por milhão em textos de notícia.

¹⁷ Esta foi uma frase produzida espontaneamente pela pesquisadora desta tese durante uma interação com sua filha de 9 anos.

consideramos a ocorrência dessa estrutura muito baixa no universo analisado e, por essa razão, não a incluímos nas demais discussões feitas abaixo.

Além das distribuições dos padrões de 1 a 5 mostradas acima, investigamos no Humanas e no C-ORAL-BRASIL outro aspecto das construções dativas do português estruturas que consideramos relevante para descrições linguísticas: o uso das preposições “a” e “para” nas construções ditransitiva preposicionada e ditransitiva preposicionada invertida. Com base nos dados da Tabela 6 e no gráfico da Figura 20, concluímos que há uma grande diferença entre os corpora escrito e oral quanto ao uso das preposições introdutórias do dativo. Enquanto a preposição “a” foi significativamente mais frequente (78,68%) do que a preposição “para” (21,32%) no Humanas, no C-ORAL-BRASIL as ocorrências da preposição “para”, incluindo formas como “pra”, “pá” e “pos”, representaram 93,47% das preposições analisadas. Na verdade, o uso da preposição “a” foi tão raro neste último (6,53%) *corpus* que é difícil postular uma alternância real entre as preposições introdutórias do dativo na fala informal da diatopia mineira do PB.

Notamos outra grande diferença entre o Humanas e o C-ORAL-BRASIL no tocante ao uso da preposição “para”. Enquanto a redução dessa preposição para “pra” e a junção da preposição com um artigo (ex: pros, pos, pruma) representam 3% das ocorrências de “para” no *corpus* escrito, no *corpus* oral essa redução e junção representam 99% das ocorrências da mesma preposição. Dados de frequência como esses, fornecidos pela análise de corpora, além de enriquecerem descrições sobre o português, podem ser muito úteis para o ensino de português como segunda língua. Afinal, ao visitarem o Brasil e interagirem oralmente com nativos, aprendizes de português/L2 se sairão bem melhor se souberem reconhecer formas como “pras”, “pos” e “pruma”, do que se souberem apenas a forma que encontram em textos escritos em português e que provavelmente terão aprendido em contextos formais de instrução, a “para”. Assim, temos um exemplo de como a Linguística de *Corpus* pode contribuir com dados para o ensino de línguas.

Tabela 6 - Distribuição das preposições “a” e “para” nos corpora do português

	Humanas	C-ORAL-BRASIL
A	78,68%	6,53%
Para	21,32%	93,47%

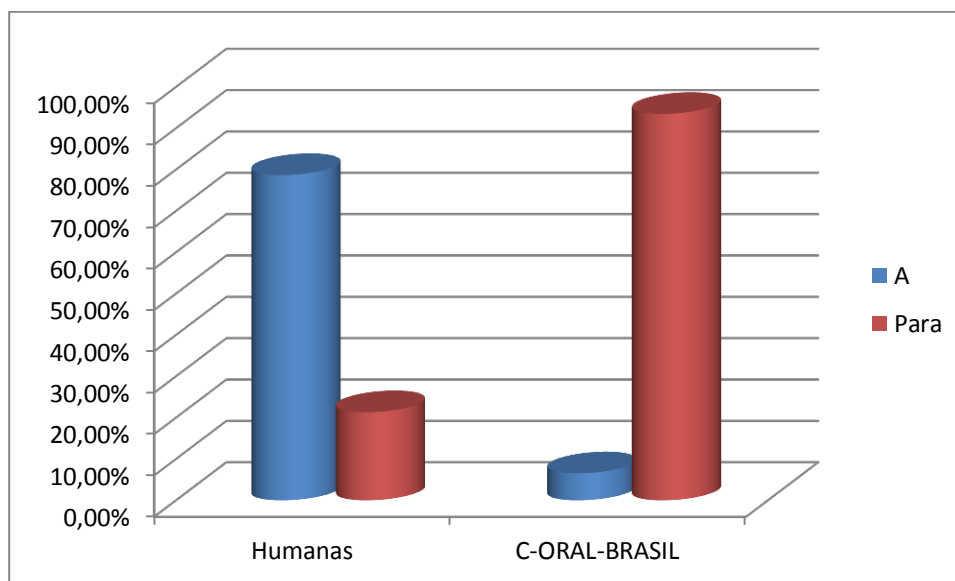


Figura 20 - Gráfico mostrando a distribuição das preposições “a” e “para” nos corpora do português.

Sugerimos que as diferenças apontadas na Figura 20 entre as preposições “a” e “para” reflitam não apenas diferentes associações entre essas preposições e os canais escrito e falado, mas também diferentes associações entre as mesmas preposições e os registros formal e informal, representados pelo Humanos e pelo C-ORAL-BRASIL, respectivamente.

Finalizamos a descrição que nos propomos a fazer sobre aspectos do português no tocante à expressão linguística de eventos de transferência de posse. Na seção 4.3 relacionamos os que já vimos sobre as frequências dos padrões frasais no português/L1 com os resultados obtidos para os padrões frasais nos corpora de falantes nativos e de aprendizes de inglês.

4.3 Distribuição das construções nos corpora de inglês L1 e L2

As frequências de construções nos corpora de inglês/L2 (*Written* e *Spoken*), da interlíngua (Br-ICLE e LINDSEI-BR) e do português/L1 (Humana e C-ORAL-BRASIL) podem ser visualizadas no gráfico da Figura 21. Como se observa, nos corpora do português/L1 destaca-se a cor vermelha, que corresponde à ditransitiva preposicionada, enquanto que nos corpora de inglês L1 e L2 predomina a cor azul, representando a ditransitiva.

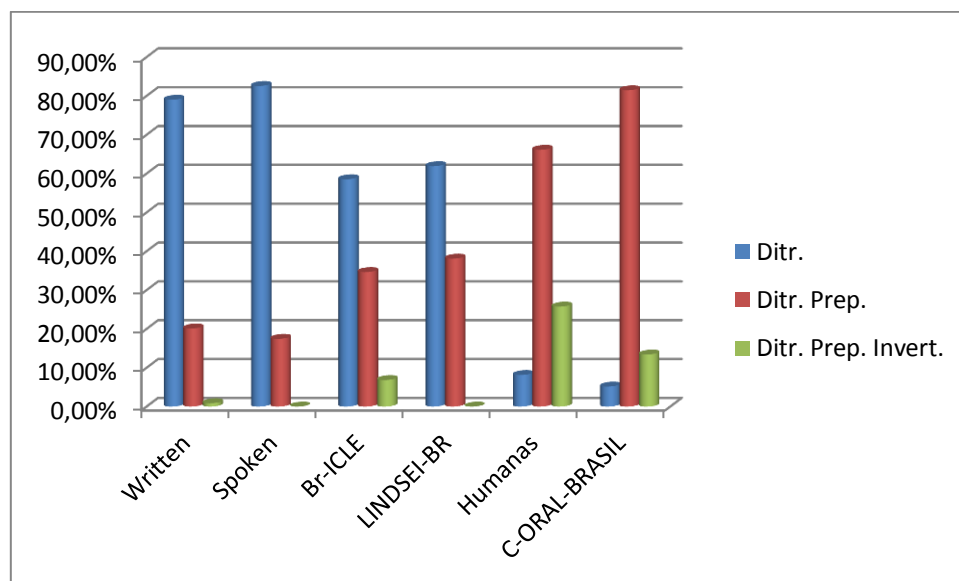


Figura 21 - Distribuição das construções nos corpora analisados.

As porcentagens seguidas das frequências brutas de cada construção por *corpus* de inglês L1 e L2 são indicadas na Tabela 7.

Tabela 7 - Distribuição das construções nos corpora de inglês/L1 e L2

	Ditransitiva Estrutura 3	Ditransitiva Preposicionada Estrutura 1	Ditransitiva Preposicionada Invertida Estrutura 2
Written	79,01% (271)	20,12% (69)	0,87% (3)
Spoken	82,56% (426)	17,44% (90)	0,00% (0)
Br-ICLE	58,54% (120)	34,63% (71)	6,83%(14)
LINDSEI-BR	61,90% (13)	38,10% (8)	0,00% (0)

Comparamos as proporções das três construções da Tabela 7 por *corpus* utilizando testes de significância estatística para confirmarmos ou rejeitarmos as hipóteses que levantamos a partir da observação do gráfico da Figura 21. Os seguintes resultados foram alcançados (Tabela 8). No *Written*, *Spoken* e *Br-ICLE*, todas as frequências são estatisticamente diferentes uma das outras.

Logo, podemos dizer que, nesses corpora, a ditransitiva (estrutura 3) é de fato mais frequente que a ditransitiva preposicionada (estrutura 1), que é mais frequente que a ditransitiva preposicionada invertida (estrutura 2). Portanto, de fato, o *corpus* escrito da interlíngua português/inglês se aproxima mais do *corpus* escrito de falantes nativos da língua alvo, com a predominância da estrutura 3, do que do *corpus* escrito da língua materna, com maior ocorrência da estrutura 1. Já no *corpus* oral da interlíngua, apesar de aparentemente a ditransitiva também ocorrer mais frequentemente, a frequência da construção ditransitiva não é estatisticamente diferente da frequência da construção ditransitiva preposicionada (p-valor = 0,2167).

Tabela 8 - Comparação entre frequências de construções por *corpus* de inglês L1 e L2 por meio da aplicação do teste exato de Fisher (*Written*, *Spoken* e LINDSEI-BR) e do qui-quadrado (Br-ICLE). Destaque para p-valor que leva à aceitação da hipótese nula.

	<i>Written</i>	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
ditransitiva X ditr. preposicionada	$\neq 1,3473\text{e-}56$	$\neq 1,0632\text{e-}104$	$\neq 2,0121\text{e-}06$	$= 0,2167$
ditransitiva X ditr. prep. invertida	$\neq 2,9006\text{e-}117$	$\neq 2,9560\text{e-}200$	$\neq 2,0642\text{e-}28$	$\neq 1,5948\text{e-}05$
ditransitiva prep. X ditransitiva prep. invertida	$\neq 7,9523\text{e-}19$	$\neq 2,2913\text{e-}29$	$\neq 8,9613\text{e-}12$	$\neq 0,0034$

$\alpha_p = 0.0166$

Embora o *corpus* escrito da interlíngua seja semelhante ao *corpus* escrito de inglês/L1 quanto à maior ocorrência em cada um deles da construção ditransitiva em relação às outras construções, o Br-ICLE se difere do *Written* quando comparamos esses dois corpora por meio do teste qui-quadrado no que diz respeito às proporções das construções dativas em cada um deles (Tabela 9). Uma vez que os p-valores obtidos indicam diferenças entre essas distribuições (p-valor = 0,004), buscamos localizar exatamente onde estavam essas diferenças. Comparamos, então, as proporções de cada construção no Br-ICLE e no *Written*, e verificamos que todas elas são diferentes: as frequências da ditransitiva, da ditransitiva preposicionada e da ditransitiva preposicionada invertida no Br-ICLE são diferentes das frequências da ditransitiva, da

ditransitiva preposicionada e da ditransitiva preposicionada invertida no *Written*, respectivamente. Olhando novamente para os dados da Tabela 7, podemos concluir, então, que no *Written* há mais construções ditransitivas do que no Br-ICLE, e que no Br-ICLE há mais construções ditransitivas preposicionadas e ditransitivas preposicionadas invertidas do que no *Written*. Todos os p-valores obtidos por meio da aplicação do teste exato de Fisher para comparar os corpora em relação às distribuições das construções são indicados na Tabela 9.

Tabela 9 - P-valores obtidos por meio da aplicação do teste exato de Fisher para comparar os corpora de inglês L1 e L2 em relação às distribuições das construções

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	= 0,0639	≠ 0,0004	= 0,1849
<i>Spoken</i>		≠ 0,0004	= 0,0344
Br-ICLE			= 0,7066

$\alpha_{\beta} = 0,0083$

Como podemos notar, os corpora de inglês L1 (*Written* e *Spoken*) foram estatisticamente semelhantes entre si (p-valor = 0,0639), e o mesmo é verdade para os corpora da interlíngua (Br-ICLE e LINDSEI-BR, p-valor = 0,1849). Portanto, nossos resultados não apontaram para uma diferença entre os canais escrito e oral no que diz respeito à distribuição das construções dativas do inglês analisadas na mesma variedade do inglês e da interlíngua, mas sim uma diferença entre os dados de inglês/L1 e inglês/L2 quanto a distribuição das construções. Essa diferença pode ser resultado da influência da L1 na interlíngua ou de diferenças entre os corpora de inglês L1 e L2 quanto às distribuições dos tipos de recipientes e temas que favorecem o uso de uma das construções dativas do inglês, como veremos na seção 4.5.2.

Focalizemos agora nossos resultados relativos à estrutura 2. A construção ditransitiva preposicionada invertida não foi observada no *corpus* oral e apresentou frequência muito baixa no *corpus* escrito de inglês/L1 (0,87%). É plausível sugerir, portanto, que essa construção ocorra raramente no inglês escrito, e mais raramente ainda no inglês oral, em contextos que não podem ser definidos a partir das 3 ocorrências dessa construção encontradas no *Written* (21a-c). Em (21c), a ditransitiva preposicionada invertida chega a ser selecionada em um contexto bastante favorável para o uso da construção ditransitiva: o verbo utilizado é fortemente associado à construção ditransitiva, o recipiente é expresso através de um pronome e o tema é

significativamente mais extenso que o recipiente.

21) Ocorrências da construção ditransitiva preposicionada invertida no *Written*:

- a) (...) lending to the houses and pedestrians a faintly unreal or even theatrical quality.
- b) (...) they issued diplomas for Breton monasteries, which conveyed to these Breton communities the same legal status and privileges as many Frankish churches enjoyed.
- c) (...) give to you an impression, which would be, I do assure you, entirely false.

A mesma tendência para a não ocorrência da estrutura 2 nos corpora de falantes de inglês/L1 foi constatada nos corpora de aprendizes de inglês: não houve nenhuma ocorrência dessa construção no *corpus* oral da interlíngua, assim como não houve no *corpus* oral de falantes nativos da língua-alvo, e sua frequência no *corpus* escrito da interlíngua foi muito baixa, também como constatamos no *corpus* escrito de inglês/L2. Não obstante, ao compararmos as frequências da ditransitiva preposicionada invertida no *Written* e no Br-ICLE, constatamos que elas são estatisticamente diferentes ($p\text{-valor} = 0,0004$). Isto é, embora tenha sido pouco frequente no Br-ICLE, a ditransitiva preposicionada invertida foi estatisticamente mais frequente nesse *corpus* do que no *Written*.

Devido à ausência da estrutura 2 nos corpora orais e sua baixa frequência nos corpora escritos de inglês L1 e L2, vamos comentar as ocorrências dessa construção separadamente na próxima seção. Em seguida, voltamos a tratar das construções ditransitiva e ditransitiva preposicionada, bastante frequentes em todos os corpora de inglês L1 e L2 por nós analisados.

4.4 Construção ditransitiva preposicionada invertida no Br-ICLE

As sentenças abaixo ilustram ocorrências da construção ditransitiva preposicionada invertida no Br-ICLE:

22) Ocorrências da construção ditransitiva preposicionada no Br-ICLE:

- a) “(...) people were walking on the street with their family, walking on the park or zoo, showing to their children the animals nature and so on.”
- b) “(...) the university needs to bring to these students a real world (...)”
- c) “(...) it’s over powering women giving to them a higher status.”
- d) “(...) they are famous for "creating" excellent professionals, offering to the student richer courses, hundreds of activities, and opportunities to do some original research.”

Analizamos primeiramente o verbo utilizado. Como se observa na Tabela 10, o segundo verbo mais frequente na construção ditransitiva preposicionada invertida foi o *bring*. O fato de esse verbo *ser* mais associado à construção ditransitiva preposicionada (GRIES & STEFANOWITSCH, 2004) pode ter favorecido a inversão de posição entre o tema e o recipiente com a preservação da preposição introdutória do dativo.

Tabela 10 - Verbos que ocorreram na construção ditransitiva preposicionada invertida no Br-ICLE

Verbo	Frequência
Show	35,7% (5)
Bring	28,6% (4)
Give	21,4% (3)
Offer	14,3% (2)

Diferentemente de *bring*, porém, os verbos *show*, *give* e *offer* favorecem o uso da construção ditransitiva (Ibidem), o que nos faz procurar outras motivações para suas ocorrências na ditransitiva preposicionada invertida. Com esse intuito, buscamos no Humanas a distribuição do verbo “mostrar”, correspondente ao verbo *show* em português, por construção (Tabela 11). Foi interessante notar que, embora no *corpus* escrito do português a construção com maior frequência total tenha sido a ditransitiva preposicionada (seção 4.3), a construção mais frequente com o verbo “mostrar” em particular nesse *corpus* foi a ditransitiva preposicionada invertida: 59,38% das ocorrências desse verbo foram na estrutura 2 *versus* 28,13% na estrutura 1 (p-valor = 0,0184). Ao observarmos esse resultado para o verbo “mostrar”, constatamos que o verbo “oferecer”, tradução do verbo do inglês *offer*, também foi mais frequente na estrutura 2 (47,2%) do que na 1 (33,04%) no Humanas (p-valor = 0,0334). Portanto, é possível que haja alguma relação entre certos verbos mais frequentes na construção ditransitiva preposicionada invertida no português/L1 e verbos que ocorrem na mesma construção na interlíngua português/inglês. Análises de amostras maiores de dados de corpora seriam necessárias para se testar essa hipótese.

Tabela 11 - Distribuição de verbos por construção no Humanas. Destaque para frequências que indicam com qual construção cada verbo foi mais frequente.

Verbo	ditransitiva preposicionada	ditransitiva preposicionada invertida	ditransitiva	pronome verbo
Dar	66,72%	15,13%	7,06%	11,09%
Trazer	51,69%	29,21%	4,49%	14,61%
Oferecer	33,04%	47,32%	4,46%	15,18%
Mostrar	28,13%	59,38%	12,50%	0,00%
Ensinar	29,17%	20,83%	16,67%	33,33%
Contar	36,36%	45,45%	0,00%	18,18%
Causar	64,00%	8,00%	16,00%	12,00%
Comprar	100,00%	0,00%	0,00%	0,00%
Passar	56,25%	37,50%	0,00%	6,25%
Entregar	46,67%	26,67%	10,00%	16,67%
Mandar	37,50%	6,25%	25,00%	31,25%
Dizer	53,85%	7,69%	23,08%	15,38%
Enviar	68,42%	21,05%	5,26%	5,26%
Perguntar	0,00%	100,00%	0,00%	0,00%
Construir	100,00%	0,00%	0,00%	0,00%

Levar	70,45%	29,55%	0,00%	0,00%
-------	--------	--------	-------	-------

Além dos verbos utilizados, analisamos nos exemplos da construção ditransitiva preposicionada invertida encontrados no Br-ICLE as distribuições das propriedades dos recipientes e temas exibidas na Tabela 12 com suas respectivas frequências e p-valores resultantes da comparação entre essas frequências.

Tabela 12 - Distribuição das propriedades dos recipientes e temas nos exemplos da construção ditransitiva preposicionada invertida no Br-ICLE e p-valores resultantes das comparações entre as frequências por coluna por meio do teste exato de Fisher. Destaque para frequências estatisticamente maiores.

	Recipiente	Tema	P-valor
Pronome	35,71% (5)	0,00% (0)	$\neq 0,0079$
NP lexical	64,29% (9)	100,00% (14)	$= 0,2329$
Animado	78,57% (11)	7,14% (1)	$\neq 0,0001$
Inanimado	21,43% (3)	92,86% (13)	$\neq 0,0010$
Dado	92,86% (13)	0,00% (0)	$\neq 1,923e-07$
Novo	7,14% (1)	100,00% (14)	$\neq 2,914e-06$
Mais extenso	0,00% (0)	92,86% (13)	$\neq 1,923e-07$

As frequências em destaque acima foram as constatadas como estatisticamente maiores, como indicam os p-valores na última coluna à direita. Note-se que recipientes pronomes, animados e dados, e temas inanimados, novos e mais extensos, destacados acima, favorecem o uso da construção ditransitiva (GRIES, 2003), ou seja, da expressão do recipiente antes do tema. Portanto, propomos que os aprendizes que produziram a construção ditransitiva preposicionada invertida no Br-ICLE foram sensíveis a diversos fatores que influenciam a ordem <recipiente tema> no inglês, porém utilizaram a preposição introdutória do dativo. Seguindo a argumentação de Pierrehumbert (2003), é possível sugerir que esses aprendizes em específico não tenham tido experiência suficiente com a língua para formarem uma sub-representação estatística do uso da construção ditransitiva preposicionada no inglês, ou seja, que eles não tenham recebido evidência negativa suficiente do insumo linguístico. Uma explicação alternativa seria que esses aprendizes

testemunharam um número maior de ocorrências da construção ditransitiva preposicionada invertida, como em “*They shared with me their deepest secrets, introduced me to their families and friends, sang to me my favorite songs and recited my favorite poetry*”¹⁸, em sua experiência com a L2 do que os demais aprendizes que contribuíram com dados para o Br-ICLE.

Voltando a nossa atenção para a influência da L1 na aquisição da L2, é interessante notar que a baixa frequência da construção ditransitiva preposicionada invertida (estrutura 2) nos corpora dos aprendizes vai ao encontro dos resultados obtidos em um contexto experimental por Zara (2009). A partir de dados de julgamento de aceitabilidade de sentenças, a pesquisadora constatou que sentenças formadas pela construção ditransitiva preposicionada invertida, como em “*Mary gave to John a book*”, tenderam a ser rejeitadas por aprendizes brasileiros dos níveis de proficiência intermediário e avançado em inglês. Portanto, concluímos que a análise de *corpus* feita nesta tese corrobora os resultados do estudo experimental de Zara no sentido de apontar para a pouca probabilidade de ocorrência de transferência da ordem <Suj. Verbo Oblíquo Objeto> do português para a interlíngua português/inglês a partir do nível intermediário de proficiência na língua-alvo. Defendemos que resultados semelhantes apontados por metodologias diferentes reforçam a confiabilidade desses resultados.

Ao falarmos da baixa probabilidade de ocorrência da construção ditransitiva invertida na interlíngua português/inglês, é necessário salientarmos que, na amostra do Br-ICLE analisada, todas as ocorrências dessa construção foram com os 4 lemas mais frequentes da amostra, os únicos de fato com uma frequência relativamente alta na amostra. Isso sugere que, em amostras maiores, com frequências mais altas de um número maior de lemas, é possível haver um número maior de ocorrências da construção ditransitiva preposicionada. Porém, a necessidade de se ter um *corpus* maior para que mais ocorrências dessa construção sejam observadas aponta para a sua infrequência relativa em comparação às demais construções, pois aspectos linguísticos mais frequentes podem ser observados em corpora com quantidades menores de textos, e aspectos linguísticos menos frequentes requerem corpora com quantidades maiores de textos para serem observados (SARDINHA, 2000b).

¹⁸ Exemplo retirado do texto “Why I’m not afraid of dying”, por David Menasche no site da CNN no dia 10 de janeiro de 2014. Disponível em http://www.cnn.com/2014/01/08/opinion/menasche-teaching-life/index.html?iid=article_sidebar. Acesso em 10 de janeiro de 2014.

Mas por que uma construção perfeitamente possível na L1 (ditransitiva preposicionada invertida / estrutura 2) foi encontrada com frequência tão baixa nos dados da interlíngua? Como vimos mais acima, no português, a estrutura 2 é significativamente menos frequente do que a estrutura 1 tanto no *corpus* escrito quanto no oral. Neste último, ela chega a ser “rara”. Sugerimos, então, que a baixa probabilidade de ocorrência desse padrão em relação a outros padrões disponíveis no português/L1 para a expressão de eventos de transferência de posse é um fator que influencie sua baixa probabilidade de transferência para a interlíngua português/inglês a partir do nível de proficiência intermediário na língua-alvo. A hipótese de efeitos de frequência no processo de transferência linguística deve ser investigada em pesquisas futuras com um número maior de construções do português.

Diferentemente da construção ditransitiva preposicionada invertida, discutida nesta seção, as construções ditransitiva e ditransitiva preposicionada foram muito frequentes em todos os corpora de inglês L1 e L2 por nós analisados. Retomemos a discussão sobre os resultados para essas construções na seção a seguir, na qual tratamos da influência que verbos particulares e que certas propriedades dos argumentos recipiente e tema exercem sobre a escolha das construções dativas do inglês.

4.5 Fatores que influenciam a escolha da construção ditransitiva e ditransitiva preposicionada nos corpora de inglês/L1 e da interlíngua português/inglês

4.5.1 Verbo

O primeiro aspecto analisado que poderia influenciar a escolha por uma das construções dativas do inglês foi o verbo utilizado nas sentenças. Abaixo mostramos os 3 lemas mais frequentes em cada *corpus* com seus respectivos números de ocorrência. Como se pode observar, o verbo mais frequente no *Spoken*, no *Written* e no Br-ICLE foi o *give*. É interessante também notar que o *corpus* oral de inglês/L1 e o *corpus* oral da interlíngua compartilham 2 entre seus 3 verbos mais frequentes: o *show* e o *tell*. Esse resultado indica que associações entre itens lexicais e o fator canal observadas no *corpus* de falantes nativos são ecoadas no *corpus* de aprendizes.

- *Spoken*: *give* (316), *show* (39), *tell* (29)
- LINDSEI-BR: *show* (9), *tell* (4), *ask* (3)
- *Written*: *give* (155), *send* (45), *offer* (15)
- BR-ICLE: *give* (101), *bring* (34), *offer* (19)

Nossa análise da frequência de verbo por construção resultou nos dados apresentados na Tabela 13. Primeiramente, deve-se notar que, em todos os corpora, poucos foram os verbos que apresentaram uma frequência mais alta, tendo a maior parte deles ocorrido com uma frequência baixa, o que dificulta o cálculo estatístico e conclusões acerca da associação entre verbos e construções. Por exemplo, é difícil saber se o verbo *make*, em destaque na Tabela 13, é mais associado à ditransitiva do que à ditransitiva preposicionada no *Written*, já que nesse *corpus* ele ocorreu apenas uma vez.

Tabela 13 - Distribuição de verbos nas construções ditransitiva (D) e ditransitiva preposicionada (DP) nos corpora de inglês/L1 e L2

<i>Written</i>									
Verbo	D	DP	D	DP	Verbo	D	DP	D	DP
Give	135	24	84,91%	15,09%	Grant	2	1	66,67%	33,33%
Send	39	6	86,67%	13,33%	Deny	2	0	100,00%	0,00%
offer	11	5	68,75%	31,25%	Deliver	1	0	100,00%	0,00%
Tell	12	1	92,31%	7,69%	Make	1	0	100,00%	0,00%
show	7	6	53,85%	46,15%	Rente	1	1	50,00%	50,00%
Wish	8	0	100,00%	0,00%	Serve	1	0	100,00%	0,00%
Hand	3	4	42,86%	57,14%	Leave	1	0	100,00%	0,00%
write	2	5	28,57%	71,43%	Afford	1	0	100,00%	0,00%
bring	5	1	83,33%	16,67%	Profit	1	0	100,00%	0,00%
cause	5	1	83,33%	16,67%	Quote	1	0	100,00%	0,00%
Buy	5	1	83,33%	16,67%	Win	1	1	50,00%	50,00%
Pay	4	2	66,67%	33,33%	Purchase	1	0	100,00%	0,00%
Lend	3	2	60,00%	40,00%	Render	1	0	100,00%	0,00%
Owe	4	0	100,00%	0,00%	Assign	0	2	0,00%	100,00%
teach	3	0	100,00%	0,00%	Get	0	4	0,00%	100,00%
promise	3	0	100,00%	0,00%	Award	0	1	0,00%	100,00%
Drop	3	0	100,00%	0,00%	Render	1	0	100,00%	0,00%
Ask	3	0	100,00%	0,00%	Deny	0	1	0,00%	100,00%
<i>Spoken</i>									
Give	264	52	83,54%	16,46%	Read	2	0	100,00%	0,00%

show	35	4	89,74%	10,26%	Make	2	0	100,00%	0,00%
Tell	29	0	100,00%	0,00%	Wish	2	0	100,00%	0,00%
Send	13	12	52,00%	48,00%	Pass	2	0	100,00%	0,00%
offer	14	6	70,00%	30,00%	Draw	2	0	100,00%	0,00%
Get	17	0	100,00%	0,00%	Charge	2	0	100,00%	0,00%
Lend	4	6	40,00%	60,00%	Design	2	0	100,00%	0,00%
Ask	9	0	100,00%	0,00%	Write	1	0	100,00%	0,00%
bring	2	2	50,00%	50,00%	Build	1	0	100,00%	0,00%
teach	4	0	100,00%	0,00%	Cut	1	0	100,00%	0,00%
Sell	1	3	25,00%	75,00%	Quote	1	0	100,00%	0,00%
leave	3	0	100,00%	0,00%	Cook	1	0	100,00%	0,00%
Hand	0	3	0,00%	100,00%	Win	1	0	100,00%	0,00%
Buy	3	0	100,00%	0,00%	Pay	1	0	100,00%	0,00%
charge	0	2	0,00%	100,00%	Cause	1	0	100,00%	0,00%
Find	2	0	100,00%	0,00%	Save	1	0	100,00%	0,00%
Save	2	0	100,00%	0,00%	Sell	1	0	100,00%	0,00%
Br-ICLE									
Give	66	32	67,35%	32,65%	Buy	2	1	66,67%	33,33%
bring	8	22	26,67%	73,33%	Hand	2	0	100,00%	0,00%
offer	10	7	58,82%	41,18%	Send	1	1	50,00%	50,00%
show	13	0	100,00%	0,00%	Ask	1	0	100,00%	0,00%
teach	8	2	80,00%	20,00%	Build	0	1	0,00%	100,00%
Tell	7	0	100,00%	0,00%	Take	0	1	0,00%	100,00%
cause	2	4	33,33%	66,67%					
LINDSEI-BR									
show	1	8	11,11%	88,89%	Write	1	0	100,00%	0,00%
Give	1	0	100,00%	0,00%	Send	1	0	100,00%	0,00%
offer	2	0	100,00%	0,00%	Ask	3	0	100,00%	0,00%
Tell	4	0	100,00%	0,00%					

Com um script do R obtivemos todos os p-valores (Tabela 14) resultantes da comparação entre as frequências de verbos na construção ditransitiva e na ditransitiva preposicionada exibidas acima. De todos os verbos que se pode visualizar na Tabela 13, os números de verbos que foram estatisticamente mais frequentes em uma das construções no *Written*, no *Spoken*, Br-ICLE e no LINDSEI-BR foram 6, 7, 5 e 2, respectivamente. Esses verbos são mostrados na Tabela 14. A partir desse resultado, concluímos que amostras maiores do que as por nós analisadas são necessárias para que possam ser notados mais claramente os diferentes graus de associação entre um número maior de verbos e as construções dativas do inglês.

Tabela 14 - P-valores resultantes da aplicação dos testes exato de Fisher e qui-quadrado para a comparação entre as frequências de verbos na construção ditransitiva e na ditransitiva preposicionada por *corpus* de inglês L1 e L2. Destaque para frequências que indicam com qual construção cada verbo ocorreu mais frequentemente.

<i>Corpus</i>	Verbo	D	DP	P-valor
<i>Written</i>	<i>Tell</i>	92,31%	7,69%	3,2690e-05
	<i>Give</i>	84,91%	15,09%	2,6940e-38
	<i>Get</i>	0,00%	100,00%	0,0285
	<i>Send</i>	86,67%	13,33%	1,3071e-12
	<i>Wish</i>	100,00%	0,00%	0,0001
	<i>Owe</i>	100,00%	0,00%	0,0285
<i>Spoken</i>	<i>Tell</i>	100,00%	0,00%	6,6517e-17
	<i>Give</i>	83,54%	16,46%	7,2304e-69
	<i>Teach</i>	100,00%	0,00%	0,0285
	<i>Show</i>	89,74%	10,26%	5,0331e-13
	<i>Get</i>	100,00%	0,00%	8,5704e-10
	<i>Ask</i>	100,00%	0,00%	4,1135e-05-
	<i>Offer</i>	70,00%	30,00%	0,0256
Br-ICLE	<i>Tell</i>	100,00%	0,00%	0,0005
	<i>Give</i>	67,35%	32,65%	1,9493e-06
	<i>Teach</i>	80,00%	20,00%	0,0230
	<i>Show</i>	100,00%	0,00%	1,9229e-07
	<i>Bring</i>	26,67%	73,33%	0,0006
LINDSEI-BR	<i>Tell</i>	100,00%	0,00%	0,0285
	<i>Show</i>	11,11%	88,89%	0,0033

Destacamos na Tabela 14 com qual construção cada verbo ocorreu mais frequentemente. Como se pode observar, o verbo *tell* foi predominante na construção ditransitiva em todos os corpora. Mais uma vez, dados de frequência de estruturas da L2 iluminam nossa compreensão sobre fatos da interlíngua. Embora brasileiros aprendizes de inglês do nível de proficiência intermediário na L2 ainda não tenham adquirido a construção ditransitiva do inglês enquanto categoria sintática abstrata, esses aprendizes demonstraram conhecimento sobre a possibilidade de o verbo *tell* em particular ser seguido por um recipiente e um tema sem o uso de uma preposição, aceitando sentenças do tipo “*I told John a story*” (ZARA, 2009). Esse aprendizado mais rápido sobre o ambiente sintático típico para a ocorrência de *tell* pode ser visto como um reflexo da sensibilidade dos aprendizes à alta frequência do verbo *tell* no padrão sintático <sujeito verbo objeto_{recipiente} objeto_{tema}> no insumo da L2, como constatamos nesta tese.

Além do resultado para o verbo *tell*, outros resultados que demonstram que aprendizes podem alcançar a mesma sensibilidade de falantes nativos da língua-alvo em relação à associação entre verbos e construções foram aqueles obtidos no *corpus* escrito de aprendizes para os verbos *give*, *teach* e *show*, associados à construção ditransitiva, e para o verbo *bring*, associado à construção ditransitiva preposicionada (GRIES & STEFANOWITSCH, 2004). Encontrar indícios de diferentes graus de associação entre construções sintáticas e construções lexicais específicas na interlíngua significa encontrar evidências de que a mente do aprendiz é capaz de capturar, de forma inconsciente, informações distribucionais probabilísticas no insumo da L2.

Devido ao pequeno número de verbos analisados no *corpus* oral de aprendizes, foi possível identificar apenas 2 verbos que exibiram frequência estatisticamente maior em uma das construções dativas do inglês. Um deles, o *tell*, foi mais frequente na ditransitiva, como esperado. O outro, o *show*, diferentemente do que constatamos no Br-ICLE, foi mais frequente na ditransitiva preposicionada, apesar de ser apontado como mais associado à ditransitiva em corpora de falantes nativos da língua-lavo (Ibidem). Reconhecemos a limitação dos nossos dados em termos quantitativos, mas podemos sugerir uma relação interessante entre esse resultado para o verbo *show* no LINDSEI-BR e outros resultados que alcançamos nesta pesquisa: o verbo *show* foi o segundo verbo mais frequente na construção ditransitiva preposicionada invertida no Br-ICLE, e o verbo “mostrar” foi mais frequente na ditransitiva preposicionada invertida do que nas demais construções no Humanas. Se, por um lado, esses resultados podem ser apenas uma coincidência, por outro lado eles podem indicar a seguinte relação: quando um verbo do português exibe maior frequência na construção ditransitiva preposicionada invertida, ele é mais fortemente associado ao uso da preposição introdutória do dativo, e seu correspondente em inglês apresenta mais chances de ocorrer na construção ditransitiva preposicionada ou na preposicionada invertida em produções da interlíngua em comparação a outros verbos.

Vimos acima que os verbos mais frequentes nos corpora analisados exibiram uma frequência estatisticamente maior em uma das construções dativas do inglês, mas que os verbos menos frequentes não. Voltando à Tabela 13, se considerarmos o total de ocorrências de cada verbo por *corpus*, perceberemos que formas do lema *give* representaram grande parte dos verbos selecionados - 46,5% dos verbos do *Written*, 61,00% dos verbos do *Spoken* e 43,9% dos verbos do Br-ICLE. Uma vez que o verbo *give* tende a interagir mais com a ditransitiva (GRIES, 2003),

a primeira hipótese plausível para explicarmos a maior frequência dessa construção nos corpora de inglês L1 e L2, conforme constatamos na seção 4.3, é a grande frequência desse lema. Porém, mesmo quando eliminamos as formas de “*give*” de nossos dados, a construção ditransitiva continua sendo a mais frequente: 74,9% no *Written*, 81,00% no *Spoken* e 51,9% no Br-ICLE. Portanto, há outros fatores que influenciam a maior frequência da construção ditransitiva nos dados analisados nesta pesquisa. Vejamos alguns desses fatores na próxima seção.

4.5.2 Propriedades dos recipientes e temas

Apresentamos nesta seção os resultados relativos a outros aspectos além do verbo utilizado que influenciam a seleção das construções dativas do inglês, a dizer, a pronominalidade, a acessibilidade discursiva, a animacidade e a extensão do recipiente e do tema. Começando pela pronominalidade, os dados da Tabela 15 revelam que, no *Written*, no *Spoken* e no LINDSEI-BR, recipientes pronomes foram estatisticamente mais frequentes que recipientes NP lexicais. Por exemplo, das 340 ocorrências de recipientes no *Written*, 61,18% são pronomes, e das 21 ocorrências de recipientes no LINDSEI-BR, 76,19% são pronomes. A maior frequência de recipientes pronomes nesses corpora, como no exemplo “*tell me all your problems*” extraído do LINDSEI-BR, ajuda a explicar a maior frequência neles da construção ditransitiva em relação à ditransitiva preposicionada (seção 4.3), pois a expressão do recipiente por meio de um pronome favorece a seleção da construção ditransitiva (GRIES, 2003; THEIJSSE, 2008). No Br-ICLE, porém, as proporções entre recipientes pronomes e recipientes NPs lexicais foram estatisticamente iguais.

Tabela 15 - Distribuição de recipientes “pronome” e “NP lexical” nos corpora de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências de recipientes pronomes e lexicais por *corpus* de inglês L1 e L2. Destaque para frequências estatisticamente maiores.

	Recipiente pronome	Recipiente NP lexical	Recipiente pronome	Recipiente NP lexical	P-valor
<i>Written</i>	208	132	61,18%	38,82%	0,0004
<i>Spoken</i>	408	108	79,07%	20,93%	0,0004
Br-ICLE	101	90	52,88%	47,12%	0,2974
LINDSEI-BR	16	5	76,19%	23,81%	0,0024

Quando comparamos os 4 corpora quanto à distribuição de recipientes pronomes e NPs lexicais por meio do teste exato de Fisher (Tabela 16), observamos diferenças estatísticas entre o *corpus* escrito e o oral de inglês/L1 (p-valor = 1,7989e-08) e entre o *corpus* escrito de aprendizes e o *corpus* oral de inglês/L1 (p-valor = 2,2228e-11). Semelhanças foram encontradas entre o *corpus* escrito de aprendizes e o *corpus* escrito de inglês/L1 (p-valor = 0,0671), e entre o *corpus* oral de aprendizes e o *corpus* oral de inglês/L1 (p-valor = 0,7848). A diferença observada entre o *Written* (escrito) e o *Spoken* (oral) e a semelhança encontrada entre o *Written* (escrito) e o Br-ICLE (escrito) apontam para uma influência do fator “canal” na distribuição de recipientes pronomes e NPs lexicais.

Tabela 16 - P-valores resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto à distribuição de recipientes pronomes e NP lexicais

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	$\neq 1,7989\text{e-}08$	= 0,0671	= 0,2461
<i>Spoken</i>		$\neq 2,2228\text{e-}11$	= 0,7848
Br-ICLE			= 0,0624

$\alpha_p = 0,0083$

Ao olharmos para a distribuição de recipientes pronomes e NP lexicais por construção, observamos uma forte associação entre “recipiente pronome” e “construção ditransitiva” em todos os corpora, como destacado na Tabela 17. No *Written* e no *Spoken*, a construção ditransitiva foi a utilizada em 92,27% e 89,46% de vezes em que um recipiente foi expresso através de um pronome, respectivamente. Nos corpora de aprendizes, a mesma tendência da seleção da construção ditransitiva quando o recipiente é expresso através de um pronome foi

identificada: de todas as ocorrências de recipientes pronomes no Br-ICLE e no LINDSEI-BR, 86,14% e 81,25% delas foram na construção ditransitiva, respectivamente.

Tabela 17 - Distribuição de recipientes pronomes por construção nos corpora de inglês/L1 e L2. Destaque para as frequências de recipientes pronomes na construção ditransitiva.

Construção	<i>Written</i>	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
Ditransitiva	92,79% (193)	89,46% (365)	86,14% (87)	81,25% (13)
Ditransitiva preposicionada	7,21% (15)	10,54% (43)	13,86% (14)	18,75% (3)

Por meio da aplicação do teste exato de Fisher, concluímos que não há diferença estatisticamente significativa entre os corpora de inglês L1 e L2 quanto à frequência de recipientes pronomes na construção ditransitiva (Tabela 18). Esse resultado indica que aprendizes são tão sensíveis quanto os falantes nativos da língua-alvo em relação à associação entre recipientes pronomes e a construção ditransitiva.

Tabela 18 - Resultado da aplicação do teste exato de Fisher para a comparação entre as frequências de recipientes pronomes que ocorrem na construção ditransitiva nos corpora de inglês L1 e L2

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	= 0,1928	= 0,0648	= 0,1255
<i>Spoken</i>		= 0,3777	= 0,3994
Br-ICLE			= 0,7014

$\alpha_{\beta} = 0,0083$

Diferenças entre os corpora foram identificadas quando consideramos os resultados para os recipientes NP lexicais. A Tabela 19 mostra as frequências desse tipo de recipiente por construção em cada *corpus* e os p-valores resultantes da comparação entre essas frequências. Enquanto no *Written* a construção ditransitiva continuou a ser a mais utilizada mesmo quando o recipiente é expresso por um NP lexical (p-valor = 0,0074), no *Spoken* não houve diferença estatisticamente significativa entre a frequência de recipientes NP lexicais na ditransitiva e na ditransitiva preposicionada, ou seja, não há indícios de que uma construção seja mais preferível que a outra na fala quando o recipiente é um NP lexical. Já nos corpora de aprendizes, ao

contrário do que ocorreu no *Written*, notamos uma associação entre recipientes NP lexicais e a construção ditransitiva preposicionada. No Br-ICLE a construção ditransitiva preposicionada foi a escolhida em 63,33% das vezes em que o recipiente foi expresso através de um NP lexical, como em “*There is no reason why you should give a gift to you partner in this date*”. No LINDSEI-BR, essa escolha ocorreu em 100% das vezes.

Tabela 19 - Frequências de recipientes Np lexicais por construção em cada *corpus* de inglês L1 e L2 e os p-valores resultantes da aplicação do teste qui-quadrado (*Written*, *Spoken* e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para comparação entre essas frequências. Destaque para frequências estatisticamente maiores.

Construção	<i>Written</i>	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
Ditransitiva	59,09% (78)	56,48% (61)	36,67% (33)	0,00% (0)
Ditr. prep.	40,91% (54)	43,52% (47)	63,33% (57)	100,00% (5)
p-valor	≠ 0,0074	= 0,0809	≠ 0,0014	≠ 0,0079

Se aprendizes tendem a associar recipientes expressos através de NPs lexicais à construção ditransitiva preposicionada, e houve uma proporção maior de recipientes NP lexicais no Br-ICLE, conforme vimos na Tabela 15, isso pode explicar porque no *corpus* escrito da interlíngua houve uma frequência maior da construção ditransitiva preposicionada em comparação ao *Written*, como vimos na seção 4.3. Em outras palavras, diferentes distribuições das propriedades dos recipientes e temas podem explicar as diferentes distribuições encontradas quando comparamos dois corpora. Essa seria uma alternativa à postulação de que a influência do português/L1 seria a responsável pelo maior número de ocorrência da ditransitiva preposicionada no *corpus* escrito da interlíngua.

Ao contrário do que observamos para a pronominalidade dos recipientes, temas expressos por NPs lexicais foram estatisticamente mais frequentes do que temas expressos por pronomes em todos os universos analisados (Tabela 20).

Tabela 20 - Distribuição de temas “pronome” e “NP lexical” nos corpora de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado (*Written*, *Spoken* e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação das frequências por coluna. Destaque para frequências estatisticamente maiores.

	Tema pronome	Tema NP lexical	P-valor
<i>Written</i>	5,29% (18)	94,71% (322)	≠ 0,0004
<i>Spoken</i>	11,24% (58)	88,76% (458)	≠ 0,0004
Br-ICLE	4,19% (8)	95,81% (183)	≠ 0,0004
LINDSEI-BR	19,05% (4)	80,95% (17)	≠ 0,0001

O fato de o tema ter sido predominantemente expresso através de um NP lexical contribui para a maior frequência geral da construção ditransitiva nos corpora de inglês L1 e L2, pois NPs lexicais favorecem a seleção dessa construção (GRIES, 2003).

Encontramos mais um dado de frequência relacionado à maior ocorrência da construção ditransitiva nos corpora de inglês L1 e L2 quando analisamos os resultados obtidos para a propriedade “animacidade”: a maior ocorrência total de recipientes animados nos corpora. Por exemplo, das 340 ocorrências de recipientes no *Written*, 86,80% são animados (ex: “he gave the young couple his blessing”) e 13,20% são inanimados, e das 191 ocorrências de recipientes no Br-ICLE, 82,40% são animados (ex: “Should we give them food”) e 17,60% são inanimados (Tabela 21).

Tabela 21 - Distribuição de recipientes animados e inanimados nos corpora de inglês L1 e L2

	Recipiente animado	Recipiente inanimado
<i>Written</i>	86,80% (294)	13,20% (46)
<i>Spoken</i>	95,16% (491)	4,84% (25)
Br-ICLE	82,72% (158)	17,28% (33)
LINDSEI-BR	100,00% (21)	0,00% (0)

As frequências de recipientes animados foram estatisticamente maiores do que as de recipientes inanimados em todos os corpora (Tabela 22).

Tabela 22 - Resultados da aplicação do teste qui-quadrado (*Written*, *Spoken* e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação entre as distribuições de recipientes animados e inanimados por *corpus* de inglês L1 e L2

	Resultado	P-valor
<i>Written</i>	Diferente	0,0004
<i>Spoken</i>	Diferente	0,0004
Br-ICLE	Diferente	0,0004
LINDSEI-BR	Diferente	3,716e-12

Não obstante, resultados de testes estatísticos¹⁹ (Tabela 23) revelam que o *Written* e o *Spoken* e o Br-ICLE e o *Spoken* diferem-se entre si quanto à distribuição dos recipientes animados e inanimados. Percebemos, assim, mais uma diferença distribucional entre os corpora escritos (*Written* e Br-ICLE) e o *corpus* oral (*Spoken*).

Tabela 23 - Resultados da aplicação do teste exato de Fisher e do teste qui-quadrado para a comparação das distribuições dos recipientes animados e inanimados nos corpora de inglês L1 e L2

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	$\neq 1,1467\text{e-}05$ (Fisher) $\neq 0,0004$ (χ^2)	$= 0,2546$ (Fisher) $= 0,2478$ (χ^2)	$= 0,0899$ (Fisher) $= 0,0849$ (χ^2)
<i>Spoken</i>		$\neq 6,0876\text{e-}07$ (Fisher) $\neq 0,0004$ (χ^2)	$= 0,6155$ (Fisher) $= 0,6156$ (χ^2)
Br-ICLE			$= 0,0505$ (Fisher) $= 0,0504$ (χ^2)

$\alpha_p = 0,0083$

Recordemos que Gries (2003) afirma que recipientes animados favorecem a escolha pela construção ditransitiva. De fato, a Tabela 24 mostra que, no *Spoken* e no *Written*, em relação ao total de ocorrências de recipientes animados, 83,30% e 82,99% interagem com a construção ditransitiva, respectivamente. Os p-valores obtidos por meio da comparação entre essas frequências e as frequências de recipientes animados na construção ditransitiva preposicionada comprovam que as diferenças observadas têm significância estatística (p-valor = 0,0004).

¹⁹ Colocamos na Tabela 23 os p-valores resultantes da aplicação do teste Exato de Fisher e do teste qui-quadrado para exemplificarmos como os resultados dos dois testes levaram às mesmas interpretações acerca da rejeição ou aceitação das hipóteses levantadas sobre os dados por nós analisados.

Similarmente, no Br-ICLE, o número de recipientes animados na ditransitiva é estatisticamente maior do que o número de recipientes animados na ditransitiva preposicionada (p-valor = 0,0004). Portanto, os aprendizes demonstram sensibilidade ao fator “animacidade” na escolha por uma das construções dativas do inglês na escrita. Já no LINDSEI-BR, não encontramos diferença estatisticamente significativa entre a frequência de recipientes animados na ditransitiva e na ditransitiva preposicionada (p-valor = 0,1994). Esse resultado obtido para o *corpus* oral de aprendizes pode indicar uma diferença entre os canais escrito e oral na interlíngua ou pode ter sido influenciado pelo menor número de dados do LINDSEI-BR. Neste último caso, resultados semelhantes aos obtidos para o Br-ICLE teriam que ser alcançadas em análise de uma amostra maior de dados orais da interlíngua.

Tabela 24 – Frequências de recipiente e temas animados e inanimados por construção nos corpora de inglês L1 e L2 e p-valores resultantes da comparação entre essas frequências por meio da aplicação do teste qui-quadrado e do teste exato de Fisher. Destaque para frequências estatisticamente maiores.

<i>Spoken</i>				
	Recip. Animado	Recip. Inanimado	Tema Animado	Tema Inanimado
Ditransitiva	83,30% (409)	68,00% (17)	1,11% (1)	99,77% (425)
Dit. Prep.	16,70% (82)	32,00% (8)	98,89% (89)	0,23% (1)
P-valor	≠ 0,0004	≠ 0,02299	≠ 0,0004	≠ 0,0004
<i>Written</i>				
	Recip. Animado	Recip. Inanimado	Tema Animado	Tema Inanimado
Ditransitiva	82,99% (244)	58,70% (27)	100,00% (2)	79,59% (269)
Dit. Prep.	17,01% (50)	41,30% (19)	0,00% (0)	20,41% (69)
P-valor	≠ 0,0004	= 0,1429	= 0,3333	≠ 0,0004
Br-ICLE				
	Recip. Animado	Recip. Inanimado	Tema Animado	Tema Inanimado
Ditransitiva	74,68% (118)	6,06% (2)	100,00% (2)	62,43% (118)
Dit. Prep.	25,32% (40)	93,94% (31)	0,00% (0)	37,57% (71)
P-valor	≠ 0,0004	≠ 7,753e-14	= 0,3333	≠ 0,0004
LINDSEI-BR				
	Recip. Animado	Recip. Inanimado	Tema Animado	Tema Inanimado
Ditransitiva	61,90% (13)	0,00%	0,00%	61,90% (13)
Dit. Prep.	38,10% (8)	0,00%	0,00%	38,10% (8)
P-valor	= 0,1994			= 0,1994

Quando comparamos os corpora de inglês L1 e L2 quanto às frequências de recipientes animados por construção (Tabela 25), chegamos aos resultados exibidos na Tabela 26, que revelam semelhanças estatísticas entre todos os corpora. Esse resultado pode ser interpretado como revelando a mesma sensibilidade à associação entre recipientes animados e a construção ditransitiva nos corpora de aprendizes e de falantes nativos da língua-alvo.

Tabela 25 - Distribuição de recipientes animados por construção nos corpora de inglês/L1 e L2

Construção	<i>Written</i>	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
Ditransitiva	82,99% (244)	83,30% (409)	74,68% (118)	61,90% (13)
Ditransitiva preposicionada	17,01% (50)	16,70% (82)	25,32% (40)	38,10% (8)

Tabela 26 - P-valores resultantes da aplicação do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes animados por construção

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	= 0,9250	= 0,0359	= 0,0249
<i>Spoken</i>		= 0,0204	= 0,0264
Br-ICLE			= 0,3008

$\alpha_p = 0,0083$

No que tange à animacidade dos temas, a Tabela 24 mostra que temas animados foram infrequentes em todos os corpora: 1 ocorrência no *Spoken*, 2 no *Written*, 2 no Br-ICLE e nenhuma no LINDSEI-BR. Consequentemente, torna-se difícil comparar os resultados para temas animados com os resultados para temas inanimados no que diz respeito à associação com as construções dativas. Já os temas inanimados, bem mais frequentes, ocorreram predominantemente na ditransitiva no *Spoken*, no *Written* e no Br-ICLE (Tabela 24).

Os próximos resultados a serem mostrados dizem respeito à propriedade “acessibilidade discursiva” dos recipientes e temas. A partir dos resultados exibidos na Tabela 27 podemos ver que os recipientes nos corpora analisados corresponderam predominantemente a informações dadas (ex: “*I’m not evening even bothering to send them postcards*”, extraído do *Written*) e os temas a informações novas (ex: “*I’d lent her some books*”, extraído do *Written*). Apenas no LINDSEI-BR o número de temas dados foi estatisticamente semelhante ao número de temas novos. Em todos os outros casos, testes estatísticos confirmam que as frequências observadas

como mais altas na Tabela 27 são estatisticamente maiores. Considerando resultados anteriores que apontam que, na construção ditransitiva, o argumento <recipiente> é dado (Gries, 2003) e o tema é não-dado (BRESNAN et al., 2007), encontramos outra motivação para a maior ocorrência dessa construção nas nossas análises dos corpora.

Tabela 27 - Distribuição de construções por recipiente e tema animados e inanimados em cada *corpus* de inglês/L1 e L2 e p-valores resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências. Destaque para frequências estatisticamente maiores.

	Recipiente dado	Recipiente novo	P-valor
<i>Spoken</i>	90,50% (467)	9,50% (49)	≠ 0,0004
<i>Written</i>	82,94% (282)	17,06% (58)	≠ 0,0004
Br-ICLE	82,72% (158)	17,28% (33)	≠ 0,0004
LINDSEI-BR	76,19% (16)	23,81% (5)	≠ 0,0016
	Tema dado	Tema novo	P-valor
<i>Spoken</i>	18,02% (93)	81,98% (423)	≠ 0,0004
<i>Written</i>	15,29% (52)	84,71% (288)	≠ 0,0004
Br-ICLE	7,85% (15)	92,15% (176)	≠ 0,0004
LINDSEI-BR	38,10% (8)	61,90% (13)	= 0,2194

Nas Tabelas 28 e 29 estão os resultados de testes estatísticos que comparam os 4 corpora no que diz respeito às proporções de recipientes dados e novos e de temas dados e novos, respectivamente. Encontramos uma diferença entre o canal escrito (*Written*) e o oral (*Spoken*) nos dados de inglês/L1 quanto à distribuição de recipientes dados e novos (p-valor = 0,0019). Mais diferenças entre os corpora foram encontradas com relação à distribuição de temas dados e novos: diferença entre o Br-ICLE e o *Spoken* (p-valor = 0,0019), entre o LINDSEI-BR e o *Spoken* (p-valor = 0,0294) e entre o LINDSEI-BR e o Br-ICLE (p-valor = 0,009). Vale ressaltar que diferenças entre os corpora no que tange às distribuições de tipos de recipientes e de pronomes considerados nesta pesquisa terá um reflexo nas proporções das construções presentes em cada *corpus*, já que cada tipo de recipiente e de pronome favorecerá o uso de uma das construções dativas do inglês.

Tabela 28 - Resultados da comparação entre as proporções de recipientes dados e novos nos corpora de inglês L1 e L2

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	$\neq 0,0019$	$= 1,0000$	$= 0,5467$
<i>Spoken</i>		$= 0,0089$	$= 0,0514$
Br-ICLE			$= 0,5627$

$\alpha_{\beta} = 0,0083$

Tabela 29 - Resultados da comparação entre as proporções de temas dados e novos nos corpora de inglês L1 e L2

	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
<i>Written</i>	$= 0,3103$	$= 0,0134$	$= 0,0164$
<i>Spoken</i>		$\neq 0,0019$	$\neq 0,0294$
Br-ICLE			$\neq 0,0009$

$\alpha_{\beta} = 0,0083$

Vejamos agora o que nossos dados mostram sobre a associação entre recipientes e temas dados e novos e as construções dativas do inglês. As frequências apresentadas na Tabela 30 sugerem, e os p-valores também exibidos nessa tabela comprovam, que tanto nos corpora de falantes nativos quanto nos corpora de aprendizes, quando o recipiente é uma informação dada, há uma preferência pela construção ditransitiva, e quando ele é uma informação nova, há uma preferência pela construção ditransitiva preposicionada. Outra semelhança entre os corpora de inglês L1 e L2 diz respeito à maior frequência de temas novos na ditransitiva.

Tabela 30 - Frequências de construções por recipientes e temas animados e inanimados em cada *corpus* de inglês/L1 e L2 e p-valores resultantes da comparação entre essas frequências por meio dos testes qui-quadrado e exato de Fisher. Destaque para frequências estatisticamente maiores.

<i>Written</i>				
	Recipiente dado	Recipiente novo	Tema dado	Tema novo
Ditransitiva	88,30% (249)	37,93% (22)	59,62% (31)	83,33% (240)
Dit. Preposicionada	11,70% (33)	62,07% (36)	40,38% (21)	16,67% (48)
P-valor	$\neq 0,0004$	$\neq 0,0164$	$= 0,0659$	$\neq 0,0004$
<i>Spoken</i>				
	Recipiente dado	Recipiente novo	Tema dado	Tema novo
Ditransitiva	87,15% (407)	38,78% (19)	49,46% (46)	89,83% (380)

Dit. Preposicionada	12,85% (60)	61,22% (30)	50,54% (47)	10,17% (43)
P-valor	≠ 0,0004	≠ 0,0399	= 1,0000	≠ 0,0004
LINDSEI-BR				
	Recip. Dado	Recip. Novo	Tema Dado	Tema Novo
Ditransitiva	81,25% (13)	0,00%	0,00%	100,00% (13)
Preposicionada	18,75% (3)	100,00% (5)	100,00% (8)	0,00%
P-valor	≠ 0,0010	≠ 0,0079	≠ 0,0001	≠ 1,923e-07
Br-ICLE				
	Recipiente dado	Recipiente novo	Tema dado	Tema novo
Ditransitiva	73,42% (116)	12,12% (4)	33,33% (5)	65,34% (115)
Dit. Preposicionada	26,58% (42)	87,88% (29)	66,67% (10)	34,66% (61)
P-valor	≠ 0,0004	≠ 4,722e-10	= 0,1431	≠ 0,0004

Observamos uma diferença entre o Br-ICLE e os corpora de inglês L1 quando comparamos os corpora quanto à distribuição de recipientes dados na ditransitiva e na ditransitiva preposicionada (Tabela 31). Uma vez constatada diferença estatisticamente significativa entre as frequências comparadas, podemos dizer que a associação entre pronomes novos e a construção ditransitiva é maior no *Written* (88,30%) e no *Spoken* (87,15%) do que no *corpus* escrito de aprendizes (73,42%).

Tabela 31 - P-valores resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto à distribuição de recipientes dados na ditransitiva e na ditransitiva preposicionada

	<i>Written</i>	Br-ICLE	LINDSEI-BR
<i>Spoken</i>	= 0,7317	≠ 0,0001	= 0,4507
<i>Written</i>		≠ 0,0001	= 0,4223
Br-ICLE			= 0,7647

$\alpha_{\beta} = 0,0083$

A última propriedade dos recipientes e temas que consideramos foi a extensão destes em número de palavras. Apresentamos na Tabela 32 as distribuições de recipientes menores que temas, de recipientes maiores que temas e de recipientes tão extensos quanto temas por *corpus*. Com base nos p-valores exibidos na Tabela 33, resultantes da comparação entre as frequências

por *corpus*, constatamos que apenas no LINDSEI-BR o número de recipientes menores que o tema não foi estatisticamente maior do que o número de recipientes maiores que temas.

Tabela 32 - Distribuições de recipientes menores que temas, de recipientes maiores que temas e de recipientes tão extensos quanto os temas por *corpus* de inglês L1 e L2. Destaque para frequências estatisticamente maiores.

	Recipiente < tema	Recipiente > tema	Recipiente = tema
<i>Written</i>	72,94% (248)	12,35% (42)	14,71% (50)
<i>Spoken</i>	75,78% (391)	6,01% (31)	18,22% (94)
Br-ICLE	63,35% (121)	17,28% (33)	19,37% (37)
LINDSEI-BR	33,33% (7)	9,52% (2)	57,14% (12)

Tabela 33 - P-valores resultantes da aplicação do teste qui-quadrado (*Written*, *Spoken*, Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para comparações entre as frequências de recipientes menores que temas, de recipientes maiores que temas e de recipientes tão extensos quanto os temas nos corpora de inglês L1 e L2. Destaque para o p-valor que leva à aceitação da hipótese nula.

	<i>Written</i>	<i>Spoken</i>	Br-ICLE	LINDSEI-BR
Recipiente < tema versus Recipiente > tema	$\neq 6,7970e-57$	$\neq 2,2266e-11$	$\neq 1,141909e-19$	$= 0,1300137$
Recipiente < tema versus Recipiente=tema	$\neq 2,3832e-52$	$\neq 4,2145e-76$	$\neq 6,5235e-18$	$= 0,2146$
Recipiente > tema versus Recipiente=tema	$= 0,4325$	$\neq 3,3137e-09$	$= 0,69153$	$\neq 0,0025$
$\alpha_{\beta} = 0,0166$				

Estudos anteriores mostram que recipientes menores que temas favorecem o uso da construção ditransitiva (Gries, 2003). Os dados obtidos para o *Written* e para o *Spoken* nesta pesquisa corroboram esses estudos, como os dados da Tabela 34 abaixo deixam evidente. Similarmente, no Br-ICLE, em relação ao total de recipientes menores que temas que identificamos no *corpus* (recipiente < tema), a construção ditransitiva foi utilizada em 85,12% das vezes (ex: “*give us reasons for living and fighting for our plans*”), uma frequência quase igual à frequência de recipientes menores que temas na ditransitiva no LINDSEI-BR, 85,71%

(ex: “offer me like a lot of courses that I could take”).

Tabela 34 – Frequências de recipientes menores, maiores e do mesmo tamanho que temas por construção nos corpora de inglês L1 e L2 e p-valores resultantes da aplicação dos testes qui-quadrado e exato de Fisher para a comparação entre essas frequências. Destaque para frequências estatisticamente maiores.

		recipiente < tema	recipiente > tema	recipiente = tema
<i>Spoken</i>	Ditransitivas	94,88% (371)	29,03% (9)	48,94% (46)
	Ditr.preposicionadas	5,12% (20)	70,97% (22)	51,06% (48)
	P-valor	≠ 0,0004	≠ 0,0014	= 0,8886
<i>Written</i>	Ditransitivas	90,32% (224)	28,57% (12)	70,00% (35)
	Ditr.preposicionadas	9,68% (24)	71,43% (30)	30,00% (15)
	P-valor	≠ 0,0004	≠ 0,0004	≠ 0,0004
<i>Br-ICLE</i>	Ditransitivas	85,12% (103)	9,09% (3)	37,84% (14)
	Ditr.preposicionadas	14,88% (18)	90,91% (30)	62,16% (23)
	P-valor	≠ < 2,2e-16	≠ 8,324e-12	= 0,06289
<i>LINDSEI-BR</i>	Ditransitivas	85,71% (6)	0,00% (0)	58,33% (7)
	Ditr.preposicionadas	14,29% (1)	100,00% (2)	41,67% (5)
	P-valor	≠ 0,02914	= 0,3333	= 0,6843

Comparamos os 4 corpora com relação às distribuições de recipientes menores que temas por construção (Tabela 35). Assim como fizemos na Tabela 23, colocamos na Tabela 35 tanto os resultados obtidos por meio da aplicação do teste exato de Fisher e quanto do qui-quadrado (χ^2) para mostrarmos como a aceitação ou rejeição da hipótese nula nos dados por nós analisados não variou segundo o teste utilizado.

Tabela 35 - P-valores resultantes da aplicação do teste exato de Fisher e do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes menores que temas por construção. Destaque para p-valores que levam à rejeição da hipótese nula.

<i>Written</i>		<i>Br-ICLE</i>		<i>LINDSEI-BR</i>	
Fisher	χ^2	Fisher	χ^2	Fisher	χ^2
<i>Spoken</i>	≠ 0,0361	≠ 0,0329	≠ 0,0010	≠ 0,0009	= 0,3177
<i>Written</i>		= 0,1627	= 0,1509	= 0,5187	= 1,0000
<i>Br-ICLE</i>				= 1,0000	= 1,0000

Diferenças foram encontradas entre os corpora oral e escrito de inglês L1 (*Spoken* e *Written*) e entre o *corpus* oral da L1 e o escrito da L2 (*Spoken* e Br-ICLE). Interpretando as diferenças entre as frequências observadas na Tabela 34 à luz dos p-valores exibidos na Tabela 35 podemos dizer que há uma maior associação entre recipientes mais curtos e a construção ditransitiva no *corpus* oral de inglês/L1 do que nos corpora escritos de inglês/L1 e da interlíngua. Desta forma, há mais um indício de diferença entre os canais oral e escrito. É estranho pensar, portanto, que não foram encontradas diferenças entre o LINDSEI-BR e os corpora escritos, mas o tamanho da amostra do *corpus* oral da interlíngua pode ser o responsável por esse resultado.

Ao contrário do que observamos para recipientes menores que temas, recipientes maiores que temas ocorreram com frequências estatisticamente maiores na construção ditransitiva preposicionada invertida no *Spoken*, no *Written* e no Br-ICLE. Apenas no LINDSEI-BR não foi constatada diferença estatística entre as frequências de recipientes maiores que temas na ditransitiva e na ditransitiva preposicionada (Tabela 34). Mais uma vez, vimos como alguns aspectos investigados nesta tese são notados apenas em corpora maiores.

Comparamos os 4 corpora em relação à distribuição de recipientes maiores que tema por construção. Inicialmente, optamos pelo teste exato de Fisher, já que a tabela feita no R para essa comparação contém 37,5% de frequências menores que 5. Podemos ver na Figura 22 os p-valores gerados, indicados entre parênteses, que nos levam à aceitação da hipótese nula, dado o nível de significância corrigido de 0,0083.

```

> tabela <-matrix(c(9,22,12,30,3,30,0,2),nc=4,byrow=FALSE)
> tabela
      [,1] [,2] [,3] [,4]
[1,]    9   12    3    0
[2,]   22   30   30    2
> colnames(tabela) <- c("Spoken","written","Br-ICLE","Lindsei-br")
> rownames(tabela) <- c("ditransitiva","ditransitiva preposicionada")
> tabela
               Spoken Written Br-ICLE Lindsei-br
ditransitiva           9     12      3         0
ditransitiva preposicionada 22    30     30        2
> comparacao_multipla_corpora(tabela,nível_significância_global=0.05,metodo="Fisher")
$Teste
[1] "Teste Exato Fisher"

$Resposta
      Written      Br-ICLE      Lindsei-br
Spoken  "= (1.0000)"  "= (0.05655597)"  "= (1.0000)"
Written ""           "= (0.04469924)"  "= (1.0000)"
Br-ICLE ""           ""                "= (1.0000)"

$Alpa
[1] 0.008333333

```

Figura 22 – Visualização dos p-valores obtidos no R resultantes da aplicação do teste exato de Fisher para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes maiores que temas por construção.

Em seguida, realizamos as mesmas comparações por meio do teste qui-quadrado. Interessantemente, as mesmas semelhanças entre os corpora foram apontadas (Figura 23), apesar do número de frequências menores que 5.

```

>
> comparacao_multipla_corpora(tabela,nível_significância_global=0.05,metodo="chisq")
$Teste
[1] "Qui-quadrado Exato"

$Resposta
      Written      Br-ICLE      Lindsei-br
Spoken  "= (1.0000)"  "= (0.05997001)"  "= (0.5852074)"
Written ""           "= (0.04547726)"  "= (0.5782109)"
Br-ICLE ""           ""                "= (1.0000)"

$Alpa
[1] 0.008333333

```

Figura 23 - Visualização dos p-valores obtidos no R resultantes da aplicação do teste qui-quadrado para a comparação entre os corpora de inglês L1 e L2 quanto às distribuições de recipientes menores que temas por construção.

Seja por meio do uso do teste exato de Fisher ou do teste qui-quadrado, concluímos,

portanto, que nossos aprendizes demonstram associar recipientes maiores que temas à construção ditransitiva preposicionada invertida da mesma forma que falantes nativos da língua-alvo.

Houve casos nos quais a extensão do recipiente e do tema em termos de número de palavras foram iguais (recipiente = tema). Quando isso aconteceu, apenas no *Written* há uma indicação de que uma das construções foi preferível: a ditransitiva.

Vimos até agora nesta seção que, considerando cada propriedade dos recipientes e temas (pronominalidade, animacidade, acessibilidade e extensão) isoladamente, nossos resultados indicam que os aprendizes são sensíveis aos mesmos fatores que influenciam a seleção das construções dativas por parte de falantes nativos da língua-alvo, mesmo que em graus diferentes. Finalmente, verificamos por meio da análise de *cluster* como os 4 corpora se aproximam ou se distanciam uns dos outros considerando simultaneamente as distribuições dessas 4 propriedades dos recipientes e temas nas duas construções dativas do inglês. A Figura 24 mostra o resultado dessa análise.

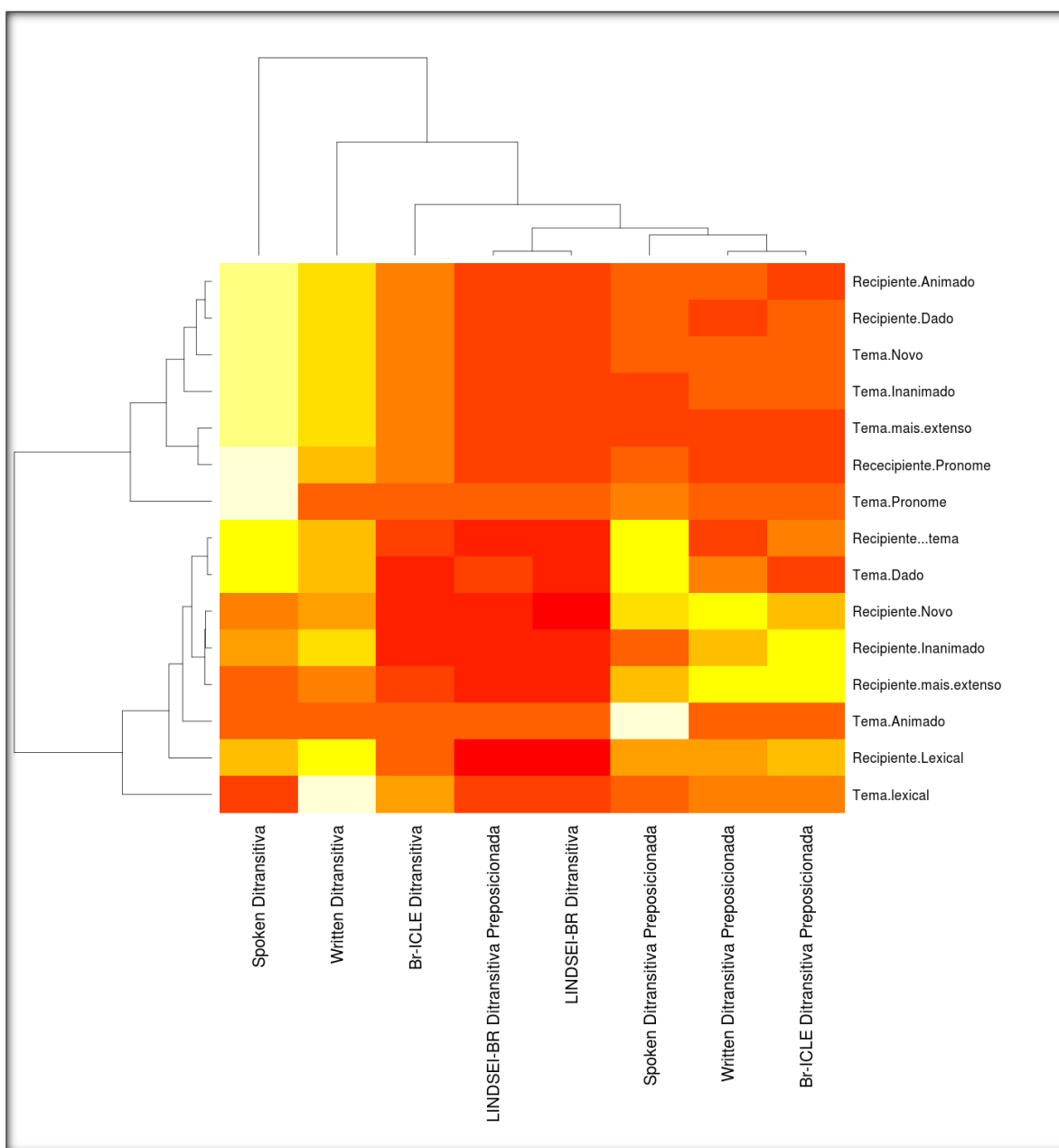


Figura 24 - Resultado da análise de *cluster* para comparar as distribuições dos diferentes tipos de recipientes e de temas que ocorrem nas construções ditransitiva e ditransitiva preposicionada nos corpora de inglês L1 e L2

Na parte superior do *heatmap* visualizamos a maior proximidade entre a ditransitiva preposicionada no *Spoken*, no *Written* e no *Br-ICLE*, e na parte mais inferior a maior proximidade entre a ditransitiva nos mesmos corpora. Isso significa que o aspecto que predominantemente determinou os agrupamentos foi a construção considerada, e não o *corpus*.

Por exemplo, a entidade “*Written* ditransitiva” foi mais próxima da entidade “Br-ICLE ditransitiva”, e não da entidade “*Written* ditransitiva preposicionada”. O fato de o *corpus Written* ter se distanciado mais de si mesmo quando foram consideradas construções diferentes, e ter se aproximado mais do *corpus* escrito da interlíngua quando a mesma construção foi considerada, mostra que falantes nativos do inglês bem como aprendizes diferenciam os usos das construções dativas do inglês.

O LINDSEI-BR foi o único *corpus* que se aproximou mais de si mesmo. Tendo em vista os resultados para os outros 3 corpora, levantamos a hipótese de que, assim como vimos em outras análises feitas nesta tese, o menor número de dados do *corpus* oral foi o responsável por esse resultado.

Ao longo deste capítulo explicamos como os resultados das análises de corpora feitos nesta pesquisa, ao mesmo tempo em que corroboram achados de estudos anteriores, contribuem com novas informações sobre o português e a interlíngua português/inglês no que tange a aspectos das construções dativas. Relacionamos a seguir, no capítulo de conclusão desta tese, esses resultados às nossas hipóteses apresentadas no capítulo 1 desta tese.

5. CONCLUSÃO

Concluimos esta tese relacionando os resultados apresentados no capítulo anterior às nossas hipóteses iniciais. Procuramos ressaltar na discussão abaixo nossas contribuições para descrições linguísticas e para a compreensão de processos da interlíngua. Por fim, traçamos perspectivas para estudos futuros.

Partimos da hipótese de que os brasileiros aprendizes dos níveis intermediário ao avançado de proficiência em inglês/L2 seriam sensíveis aos fatores que influenciam a escolha por uma das construções dativas do inglês, e essa hipótese foi confirmada. Por exemplo, nossos dados indicam que, no *corpus* escrito de aprendizes, há uma maior associação entre *tell, give, teach* e *show* e a construção ditransitiva, assim como observado em corpora de falantes nativos de inglês (GRIES & STEFANOWITSCH, 2004). Além disso, constatamos que nossos aprendizes são tão sensíveis quanto os falantes nativos da língua-alvo em relação à associação entre recipientes pronomes e animados e a construção ditransitiva do inglês (GRIES, 2003). Considerando as 4 propriedades dos recipientes e temas que analisamos simultaneamente, ficou claro na análise de *cluster* que brasileiros aprendizes de inglês diferenciam na escrita os contextos preferenciais de uso da construção ditransitiva e da ditransitiva preposicionada do inglês. Destacamos a relevância desse tipo de conhecimento para que se alcance proficiência em uma língua (WULFF & GRIES, 2006). Além disso, propomos que nossos resultados relativos à sensibilidade que aprendizes têm aos fatores que influencia a escolha pelas construções dativas do inglês indicam que brasileiros aprendizes de inglês são capazes de capturar informações probabilísticas do insumo linguístico que provavelmente²⁰ não foram apresentadas a eles por meio de instrução formal.

Em alguns casos, encontramos diferenças entre os corpora de inglês L1 e L2. Por exemplo, observamos no Br-ICLE uma associação maior entre recipientes NP lexicais e a construção ditransitiva preposicionada do que no *Written*. Verificamos também que a associação entre recipientes novos e a construção ditransitiva é maior no *Written* e no *Spoken* do que no *corpus* escrito de aprendizes. Diferenças como essas entre os corpora de falantes nativos e de

²⁰ Em 20 anos de experiência com a língua inglesa, nunca testemunhei o ensino sobre os contextos preferenciais de uso das construções dativas do inglês.

aprendizes de inglês não são surpreendentes, pois mesmo quando corpora de falantes nativos de inglês que pertencem a grupos diferentes em termos de idade, gênero e variedade do inglês são comparados, diferenças entre o grau de importância das variáveis que influenciam a seleção das construções são encontradas (BRESNAN & HAY, 2008; KENDALL et al., 2011; THEIJSEN et al., s/d). em outras palavras, considerando que os falantes/escritores que contribuíram com dados para os corpora de inglês L1 e L2 analisados nesta tese pertencem a grupos diferentes em diversos aspectos, podemos dizer que as diferenças encontradas em nosso estudo são esperadas com base em estudos anteriores com diferentes variedades de inglês/L1.

Nossas hipóteses sobre a L1 dos aprendizes também foram confirmadas: a construção ditransitiva preposicionada é a configuração mais frequente para a expressão de eventos de transferência de posse e a omissão da preposição é rara. Além de confirmar essas hipóteses baseadas em estudos anteriores, nossa pesquisa contribui com um dado quantitativo relevante: a baixa ocorrência da construção ditransitiva preposicionada invertida em relação à construção ditransitiva preposicionada, especialmente na fala. Assim, embora sentenças com a estrutura <sujeito verbo oblíquo objeto> sejam possíveis no português, a probabilidade de que elas ocorram é significativamente menor do que a probabilidade de ocorrência de frases com a ordem <sujeito verbo objeto oblíquo> na escrita culta, e é ainda menor na fala informal. Vimos também que, nas construções dativas do português, a preposição “a” é mais frequente na escrita formal, e a preposição “para”, em sua forma reduzida e unida a pronomes (pra, pos etc.), é mais frequente na fala informal. Concluímos, portanto, que as variantes estruturais do português do Brasil consideradas nesta tese devem ser diferenciadas quanto às suas frequências de ocorrência e suas associações com diferentes canais e registros. Assim, distinguimos o que é linguisticamente possível do que é linguisticamente provável (CHATER & MANNING, 2006) em contextos específicos de uso.

O parágrafo acima já mostra que foi confirmada a nossa hipótese de que haveria alguma diferença entre os corpora escritos e orais quanto a aspectos das construções investigadas. Outros dados que confirmam essa hipótese incluem diferenças entre os corpora escrito e oral de inglês/L1 quanto às distribuições de recipientes pronomes e NPs lexicais e quanto ao uso de NPs lexicais na construção ditransitiva.

Finalmente, nossa hipótese de que a transferência do padrão sintático <Sujeito Verbo Oblíquo Objeto> do português para a interlíngua poderia ocorrer com baixa frequência foi

confirmada pelos dados do Br-ICLE, mas não do LINDSEI-BR, já que neste último não encontramos ocorrências dessa estrutura. Essa diferença entre os dois corpora da interlíngua analisados pode refletir mais uma diferença entre a escrita e a fala ou podem ser resultado do menor número de sentenças extraídas do LINDSEI-BR.

Os dados de frequência da L1, da L2 e da interlíngua obtidos nesta tese, juntamente com resultados de estudo anterior sobre a aquisição das construções dativas do inglês por aprendizes brasileiros (ZARA, 2009), permitiu-nos sugerir as seguintes relações para a aquisição de construções da estrutura argumental da L2 e para a influência da L1 nesse processo. Construções frequentes na L1 e na L2, como no caso da construção ditransitiva preposicionada no português e no inglês, são adquiridas mais rapidamente por aprendizes iniciantes. Construções presentes na L2, como a ditransitiva no inglês, mas com ocorrência em contextos restritos de algumas variedades da L1, são adquiridas apenas nos estágios avançados de proficiência na L2. Construções possíveis na L1, mas com baixa frequência de ocorrência em relação à outra variante estrutural, como no caso da ditransitiva preposicionada do português, tendem a não ser transferidas para a interlíngua nos estágio intermediário e avançado de proficiência na L2. Padrões sintáticos frequentes na L1, como a ordem <Pronome_{recipiente} Verbo Objeto_{tema}> no português, mas não testemunhadas no insumo da L2, não são transferidos para a interlíngua. Esses efeitos de frequência por nós propostos são resumidos na Figura 25.

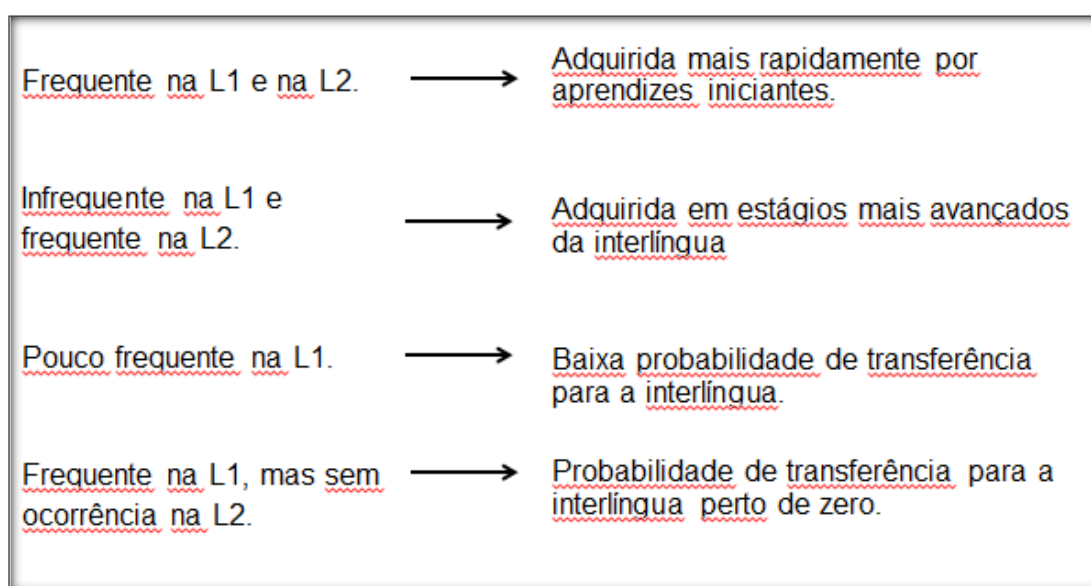


Figura 25 - Resumo das relações sugeridas entre efeitos de frequência da L1 e da L2 em processos da interlíngua

Tanto as nossas interpretações sobre a influência do português/L1 na interlíngua português/inglês quanto os nossos resultados sobre o conhecimento que brasileiros aprendizes de inglês possuem acerca das construções dativas do inglês sugerem que os elementos básicos da teoria de probabilidades, os efeitos de frequência, atuam na interlíngua assim como no sistema linguístico que é o conhecimento sobre a L1 (BOD, HAY & JANNEDY, 2003). Portanto, ao olharmos para a interlíngua, podemos contribuir com a formulação de teorias linguísticas.

Procuramos destacar neste capítulo de conclusão como nossa pesquisa contribui com descrições sobre o português do Brasil e sobre a interlíngua português/inglês, e com discussões acerca de fenômenos da interlíngua. Tudo isso com a vantagem da análise de *corpus* que é a observação de dados mais generalizáveis que refletem o uso da língua em suas diferentes modalidades. Devemos reconhecer, porém, que o tamanho da amostra do LINDSEI-BR por nós analisada foi uma limitação de nosso estudo, e que estudos futuros devem considerar amostras maiores da interlíngua português/inglês falada.

Outra sugestão para pesquisas futuras é a análise de dados de aprendizes brasileiros de níveis de proficiência mais iniciantes, a partir da qual seria possível verificar, entre outros, se a influência da L1 na interlíngua ocorre com maior frequência em estágios mais iniciais de aquisição da L2. Compilar um *corpus* representativo de brasileiros aprendizes iniciantes de inglês é um desafio que deve ser aceito por pesquisadores que investigam o desenvolvimento da interlíngua.

Estudos futuros podem ainda lançar mão de modelos probabilísticos que revelem mais sobre como as construções dativas do inglês são escolhidas em produções da interlíngua. Cada vez mais, o domínio de técnicas estatísticas mostra-se necessário para o avanço de estudos linguísticos.

Enfim, muito ainda se pode desvendar sobre as línguas de falantes nativos e de aprendizes a partir da Linguística de *Corpus* e da Linguística Probabilística. Esperamos que esta tese tenha exemplificado como essas abordagens para estudos linguísticos são promissoras em revelar aspectos sobre a fascinante linguagem humana.

REFERÊNCIAS

ALUÍSIO, S.M.; OLIVEIRA, L.H.M.; PINHEIRO, G. M. Os tipos de anotações, a codificação, e as interfaces do Projeto Lácio-Web: quão longe estamos dos padrões internacionais para *corpus*? In: Workshop em Tecnologia da Informação e da Linguagem Humana, Salvador, Centro de Convenções. **Anais do II TIL**. 5 a 6 de agosto de 2004 Disponível em: <<http://www.nilc.icmc.usp.br/lacioweb/publicacoes.htm>>. Acesso em: 11 de abril de 2013.

ARMELIN, P. R. G. Sentenças bitransitivas do português brasileiro: aspectos informacionais, semânticos e sintáticos. In: **Estudos Linguísticos**, São Paulo, 39 (1): p. 419-432, v. 39, n. 1, 2010. Disponível em: <http://www.gel.org.br/estudoslinguisticos/volumes/39/v2/el_v39n1_32.pdf>. Acesso em: 26 de abril de 2012.

ARPPE, A; GILQUIN, G.; GLYNN, D.; HILPERT, M.; ZESCHEL, A. Cognitive *Corpus* Linguistics: five points of debate on current theory and methodology. **Corpora**, 5/2, 2010.

AYRES-BENNETT, W. Sociolinguistic variation in seventeenth-century France. Methodology and case studies. Cambridge: Cambridge University Press, 2004.

BAAYEN, R. H. **Analyzing linguistic data**: a practical introduction to statistics using R. Cambridge University Press, 2008.

BARONI, M.; EVERT, S. Statistical methods for *corpus* exploitation. In: Anke Lüdeling and Merja Kytö (Eds.), **Corpus linguistics**: An international handbook, Volume 2, Berlin: Mouton de Gruyter: p. 777-802, 2009.

BARRON, ANNE. **Acquisition in interlanguage pragmatics**: learning how to do things with words in a study abroad context. Amsterdam; Philadelphia, PA: J. Benjamins Pub. CO., 2003. 398 p.

BENSON, C. Key concepts in ELT: transfer/cross-linguistic influence. **ELT JOURNAL**, 56/1: p. 68-70, 2002.

BIBER, D. Representativeness in *corpus* design. Literary and linguistic computing, Volume 8,

Oxford: Oxford University Press: p. 243-257, 1993.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics**: investigating language structure and use. Cambridge; New York: Cambridge University Press, 1998.

BISPO, K. C. I. F. A Sintaxe do Dativo no Português. In: Cadernos do CNLF, vol. VIII, nº 14, Questões de Morfossintaxe. Rio de Janeiro: Instituto de Letras da UERJ, 2004.

BISPO, K. C. I. F.; SALLES, H. M. M. L. **Estudo comparativo do dativo no Português Brasileiro e em Línguas Românicas e Germânicas**. Universidade de Brasília, 2005. Disponível em: www.unicamp.br/.../um-estudocomparativo-do-dativo-1483. Pdf. Acesso em: 01 de agosto de 2011.

BLEY-VROMAN, R.; JOO, H. The acquisition and interpretation of english locative construction by native speakers of Korean. **Studies in Second Language Acquisition**, 23, p. 207-219, 2001.

BOAS, H. C. Cognitive construction grammar. In: HOFFMANN, T.; TROUSDALE, G. (Eds.), **The Oxford Handbook of construction grammar**. Oxford: Oxford University Press, 2013. p. 233-254.

_____. Towards a lexical-constructional account of the Locative Alternation. In: L., CARMICHAEL, C.-H. HUANG AND V. SAMIAN (Eds.), **Proceedings of the 2001 Western Conference on Linguistics**, 13, 2003. p. 27-42.

BOD, R. Introduction to elementary probability theory and formal stochastic language theory. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds). **Probabilistic linguistics**. Cambridge, Mass.: The MIT Press, 2003.

BOD, R.; HAY, J.; JANNEDY, S. Introduction. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds). **Probabilistic linguistics**. Cambridge, Mass.: The MIT Press, 2003.

BRESNAN, J.; CUENI, A.; NIKITINA, T.; BAAYEN, R. H.. Predicting the Dative Alternation.” In: G. BOUMA, I. KRAEMER AND J. ZWARTS (Eds). **Cognitive foundations of interpretation**. Royal Netherlands Academy of Science, Amsterdam, 2007. p. 69-94.

BRESNAN, J.; FORD, M. **Predicting syntax**: processing dative constructions in American and

Australian varieties of English. *Language*. 86(1), 2010. p. 186-213.

BRESNAN, J.; HAY, J. Gradient Grammar: an effect of animacy on the syntax of give in New Zealand and American English. *Lingua* 118: 245-259. 2008.

British National *Corpus*. University of Oxford, 2010. Disponível em: <http://www.natcorp.ox.ac.uk>. Acesso em: 06 de julho de 2011.

BROOKE, J.; HIRST, G. Measuring interlanguage: native language identification with l1-influence metrics. In: CALZOLARI, N., CHOUKRI, K., DECLERCK, T., DOGAN, M. U., MAEGAARD, B., MARIANI, J., ODIJK, J., AND PIPERIDIS, S., (Eds). **Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)**, Istanbul, Turkey, 2012. p. 779–784.

BROWN, H. D. **Teaching by principles**: an interactive approach to language pedagogy. 3. ed. Pearson ESL, 2007.

CALLIES, M.; SZCZESNIAK, K. Argument realization, information status and syntactic weight: a Learner-*Corpus* Study of the Dative Alternation. In: GROMMES, P. AND M. WALTER (Eds.). *Fortgeschrittene Lernervarietäten. Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer, 2008. p. 165-187.

CAMPBELL, A.; TOMASELLO, M. The acquisition of English dative constructions. In: **Applied Psycholinguistics**, vol. 22, 2001. p. 253–267.

CANÇADO, M. **Manual de semântica**: noções básicas e exercícios. BH: UFMG, 2005.

CANTOS GOMEZ, Pascual. **Do we need statistics when we have linguistics?** DELTA [online]. 2002, vol.18, n.2, pp. 233-271. ISSN 0102-4450.

CAVALCANTE, R. Complementos dativos sem preposição no dialeto mineiro. *ReVEL*, vol. 7, n.12, 2009.

CHATER, N.; MANNING, C. D. Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7), 2006. p. 335-344.

COOPER, T. Teaching idioms. *Foreign Language Annals*, 31, 1998. p. 255-266.

CORDER, S. P. A Role for the Mother Tongue. In: GASS, S.; SELINKER, L. (Orgs.) **Language transfer in language learning**. Amsterdam: John Benjamins, 1993. p.18-31.

_____. Language distance and the magnitude of the language learning task. *Studies in Second Language Acquisition*, vol. 2, p. 27–36, 1978.

CORDER, G. W. & FOREMAN, D. I. *Nonparametric Statistics for Non-Statisticians: a Step-by-Step Approach*. New Jersey: Wiley, 2009. Capítulo 8. p. 155-191.

CORPUS DO PORTUGUÊS. Disponível em <http://www.corpusdoportugues.org/x.asp>. Acesso em: 16 de maio de 2011.

COSTA, S. F. **Introdução ilustrada à estatística**. 2. Ed. São Paulo: HARBRA, 1992.

CROFT, W.; CRUSE, D. A. **Cognitive linguistics**. Cambridge, U.K., New York: Cambridge University Press, 2004. 356p.

CUERVO, M. C. Double objects in Spanish as a second language: acquisition of morphosyntax and semantics. *Studies in Second Language Acquisition*, 29, 2007. p. 583-615.

FURTADO DA CUNHA, M. A. A construção ditransitiva do português do Brasil. In: *Anais do VII Congresso Internacional da ABRALIN*, Curitiba, 2011, Curitiba.

DAVIES, W. D. English dative alternation and evidence for a Thematic Strategy in Adult SLA. *Issues in Applied Linguistics*, 5(1), Department of Applied Linguistics, UC Los Angeles, 1994.

ELLIS, N. C. Constructions, chunking, and connectionism: the emergence of second language structure. In: DOUGHTY, C. J.; LONG, M. H. (Eds.) **The handbook of second language acquisition**. Oxford: Blackwell, 2003. p. 63-103.

_____. Frequency effects in language processing. *Studies in Second Language Acquisition*, 24, 2002. p. 143–188.

ELLIS, N. C.; LARSEN-FREEMAN, D. Constructing a second language: analyses and computational simulations of the emergence of linguistic constructions from usage. **Language Learning**, v. 59, 2009. p. 93-128.

ELLIS, N. C.; FERREIRA-JUNIOR, Fernando. Construction learning as a function of frequency, frequency distribution, and function. **Modern Language Journal**, v. 93, 2009. p. 370-385.

ELLIS, R. Item versus system learning: explaining free variation. **Applied Linguistics**, vol. 20, 1999. p. 460-480

_____. **Second language acquisition**. Oxford University Press, Oxford, 1997.

_____. **The study of second language acquisition**. Oxford: Oxford University Press, 1994. 824 p.

_____. **Understanding second language acquisition**. Oxford: Oxford University Press, 1985. 327 p.

ELLIS, R.; BARKHUIZEN, G. P. **Analysing learner language**. Oxford: Oxford University Press, 2005.

EVISON, J. M. What are the basics of analysing a *corpus*?. In: MCCARTHY, M. J.; O'KEEFFE, A. (Ed.). **The Routledge Handbook of Corpus Linguistics** Routledge, 2010. p. 122-135.

FEW, S. Multivariate Analysis Using Heatmaps. **Visual Business Intelligence Newsletter**, 2006. Disponível em <http://www.perceptualedge.com/articles/b-eye/heatmaps.pdf>. Acesso em: 12 de janeiro de 2014.

FRISHKOFF, G., LEVIN, L., PAVLIK, P., IDEMARU, K., JONG, N. A model-based approach to second-language learning of grammatical constructions. In: **Proceedings of the 30th conference of the cognitive science society**, Washington, D.C., 2008. p. 916–921

GABAS JR., N. Linguística histórica. In: MUSSALIM, Fernanda; BENTES, Anna Christina. (Orgs.). **Introdução à linguística: domínios e fronteiras**. São Paulo: Cortez, 2001, p. 77-103.

GASS, S.; SELINKER, L. **Second language acquisition: an introductory course**. Hillsdale: L. Erlbaum, 1994. 357p.

GIBBS, R.W. What do idioms really mean? **Journal of Memory and Language**, 31, p. 485-506,

1992.

GOLDBERG, A. E. Verbs, constructions and semantic frames. In: M. RAPPAPORT HOVAV, M; DORON, E; SICHEL, I. (Ed.). **Syntax, lexical semantics and event structure**. Oxford University Press, 2010. p. 39-58.

_____. **Constructions at work**: the nature of generalization in language. Oxford: Oxford University Press, 2006. 280p.

_____. Constructions: a new theoretical approach to language. **Trends in Cognitive Science**, vol. 7, p. 219-224, 2003.

_____. **Constructions**: a construction grammar approach to argument structure. Chicago: University of Chicago Press, 1995. 241p.

GOLDBERG, A.; CASENHISER, D.; SETHURAMAN, N. A lexically based proposal of argument structure meaning. **Chicago Linguistic Society**, vol 39, p. 67-81, 2003.

GOLDBERG, A. E.; CASENHISER, D.; WHITE, T. R.. Constructions as categories of language. **New Ideas in Psychology**, vol. 25, p. 70-86, 2007.

GOMES, C. A. Dative alternation in Brazilian Portuguese: typology and constraints. **Language Design Journal of Theoretical And Experimental Linguistics**. Universidade de Granada, v. 5, p. 67-78, 2003.

GONZÁLEZ- ROMERO, L. On the use of the English dative alternation by Spanish learners of English. A *corpus*-based study. **The 44th Annual BAAL Conference**, University of the West of England, 1-3 September 2011

GRIES, S. T. **Statistics for linguistics with R**. 2nd rev. and ext. ed. Berlin & New York: De Gruyter Mouton, 2013.

_____. **Quantitative corpus linguistics with R**: a practical introduction. New York: Routledge, 2009a.

_____. What is *corpus* linguistics? **Language and Linguistics Compass**, 3, p. 1-17, 2009b.

_____. "Towards a *corpus*-based identification of prototypical instances of constructions." **Annual Review of Cognitive Linguistics** . v. 1 p. 1–27. 2003.

_____. A multifactorial analysis of syntactic variation: particle movement revisited. **Journal of Quantitative Linguistics**, 8(1). 33-50, 2001.

GRIES, S. T.; STEFANOWITSCH, A. Channel and constructional meaning: a collostructional case study. In: KRISTIANSEN, G.; DIRVEN, R. (Eds). **Cognitive sociolinguistics: language variation, cultural models, social systems**. Berlin and New York: Mouton de Gruyter, 2008. p. 129–152.

_____. Extending collostructional analysis: a *corpus*-based perspective on 'alternations'. **International Journal of Corpus Linguistics**, 9.1:97-129, 2004.

GRIES, S. T.; WULFF, S. Do foreign language learners also have constructions? Evidence from priming, sorting, and corpora. **Annual Review of Cognitive Linguistics**, 3, 2005. p. 182-200.

GRIMSHAW, J. B. **Argument structure**. Cambridge, London: MIT, 1992. 201p.

GROEFSEMA, M. The real-world colour of the dative alternation. **Language Sciences**, 23/4, p. 525-550, 2001.

HAMDAN, J. **Language transfer and the acquisition of the english dative alternation by native speakers of Arabic**. 1994. (PhD dissertation) - University of Reading, UK, Disponível em http://jujo.academia.edu/JihadHamdan/Papers/906107/Language_Transfer_and_Aquisition_of_the_English_Dative_Alternation_by_Native_Speakers_of_Arabic. Acesso em: 03 de abril de 2012.

HOPP, H. Ultimate attainment in L2 inflection: performance similarities between non-native and native speakers. **Lingua**, 120(4), p. 901-931, 2010.

HOWELL, D. C. Chi-square test: analysis of contingency tables. In: **Lovric, M. International Encyclopedia of Statistical Science**, Springer-Verlag, Berlin, 2011.

INAGAKI, S. Japanese and chinese learners' acquisition of the narrow range rules for the dative alternation in English. **Language Learning**, 47, p. 637-669, 1997.

INTERNATIONAL *CORPUS* OF LEARNER ENGLISH. Disponível em:
<http://www.uclouvain.be/en-cecl-icle.html>. Acesso em 10 de maio de 2011.

INTERNATIONAL *CORPUS* OF ENGLISH – BRITISH COMPONENT. Disponível em:
<<http://www.ucl.ac.uk/english-usage/projects/ice-gb/>>. Acesso em 16 de agosto de 2011.

JARVIS, S. Conceptual transfer: crosslinguistic effects in categorization and construal. **Bilingualism: Language and Cognition**, 14, 1-8, 2011.

_____. Methodological rigor in the study of transfer: identifying L1 influence in the interlanguage lexicon. **Language Learning**, 50, no. 2, p. 245-309, 2000.

JOO, H. R. Second language learnability and the acquisition of the argument structure of English locative verbs by Korean speakers. **Second Language Research**, 19, 4, p. 305–328, 2003.

JURAFSKY, D. Probabilistic modeling in Psycholinguistics: linguistic comprehension and production. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds). **Probabilistic linguistics**. Cambridge, Mass.: The MIT Press. 2003.

KANG, S. The Acquisition of english dative constructions by Korean EFL Children. **Proceedings of The 16th Conference of Pan-Pacific Association of Applied Linguistics**, 2011. The Chinese University of Hong Kong, 2011. Disponível em
<http://www.paaljapan.org/conference2011/ProcNewest2011/pdf/oral/1F-1.pdf>. Acesso em: 09 de abril de 2012.

KELLERMAN, E. Now you see it, now you don't. In: GASS, S.; SELINKER, L. (Eds.), **Language transfer in language learning**. Rowley, MA: Newbury House, 1983. p. 112-134

_____. Transfer and non-transfer: where we are now. **Studies in Second Language Acquisition**, 2, p. 37-57, 1979.

KEMMERER, D. Action verbs, argument structure constructions, and the mirror neuron system. In: M. Arbib (Ed.). **Action to language via the mirror neuron system**. Cambridge University Press. 2006.

KENDALL, T.; BRESNAN, J.; HERK, G. V. The dative alternation in African American English: Researching Syntactic Variation and Change across Sociolinguistic Datasets. **Corpus linguistics and linguistic theory**, 7.2, p. 229-244, 2011.

KENNEDY, G. D. **An introduction to corpus linguistics**. London; New York: Longman, 1998.

KRAMSCH, C.; WHITESIDE, A. Three fundamental concepts in second language acquisition and their relevance in multilingual contexts. **The Modern Language Journal**, 91: 907–922, 2007.

UFMG. Laboratório de Estudos Empíricos e Experimentais da Linguagem. **Projetos em andamento**. Disponível em <http://www.lettras.ufmg.br/CMS/index.asp?pasta=leel&path=2010122910448.asp&title=Projetos%20em%20andamento>. Acesso em 25 de maio de 2011.

LAKOFF, G. The contemporary theory of metaphor. In: ORTONY, A. **Metaphor and Thought**. Second Edition: Cambridge University Press, 1993. p. 202–252.

_____. **Women, fire, and dangerous things**: what categories reveal about the mind. Chicago: editor, 1987. 614p

LAKOFF, G. Metametaphorical Issues: A Figure of Thought. In *Metaphor and Symbolic Activity*. Vol. I (3). p.215-225, 1986.

LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. Chicago: University of Chicago Press, 1980.

LAKSHMANAN, U.; SELINKER, L. Analysing interlanguage: how do we know what learners know? **Second Language Research**, 17, p. 393-420, 2001.

LANGACKER, R. W. **Foundations of cognitive grammar**. Stanford: Stanford University Press, vol 1, 1987.

LEVIN, B. **English verb classes and alternations**: a preliminary investigation. University of Chicago Press, Chicago, IL, 1993.

LINDSEI. Disponível em <http://www.uclouvain.be/en-cecl-lindsei.html>. Acesso em 15 de maio de 2011.

LITTAUER, R. **Visualising language typology** – Plotting WALS with Heat Maps. 2012. Página

da web: <http://www.replicatedtypo.com/visualising-language-typology-plotting-wals-with-heat-maps/5189.html>. Acesso em 08 de janeiro de 2014.

LUCCHESI, D. & MELLO, C. A alternância dativa no português afro-brasileiro: um processo de reestruturação original da gramática. **PAPIA**, 19, p. 153-184, 2009.

MANNING, C. D. Probabilistic syntax. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds). **Probabilistic linguistics**. Cambridge, Mass.: The MIT Press, 2003.

MATEUS, M. H. M.; BRITO, A. M.; DUART, I; FARIA, I. H. **Gramática da língua portuguesa**. Lisboa: Caminho, 2003, 5ª edição.

MAZURKEWICH, T. The acquisition of the dative alternation by second language learners and linguistic theory. **Language Learning**, 34: 91-109, 1984.

MCDONALD, J.H. **Handbook of biological statistics**. 2nd ed.. Sparky House Publishing, Baltimore, Maryland, 2009. Disponível em: <http://udel.edu/~mcdonald/statFishers.html>. Acesso em 02 de dezembro de 2013.

MCENERY, T.; HARDIE, A. **Corpus linguistics: method, theory and practice**. Cambridge: Cambridge University Press, 2012.

MCENERY, T.; XIAO, R.; TONO, Y. **Corpus-based language studies: an advanced resource book**. London; New York: Routledge, 2006.

MCENERY, T.; WILSON, A. **Corpus linguistics: an introduction**. Edinburgo: Edinburgh University Press, 2001.

MCNEIL JR., D. G. **Bill Gates Donates \$750 Million to Shore Up Disease-Fighting Fund**. Disponível em http://www.nytimes.com/2012/01/27/health/gates-donates-750-million-to-global-fund.html?_r=1&. Acesso em: 10 de janeiro de 2014.

MELLO, H. R. Os corpora orais e o C-ORAL-BRASIL. In: RASO, Tommaso; MELLO, Heliana. (Orgs.). **C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal**. Belo Horizonte: Editora UFMG, 2012, p. 31-54.

MELLO, H. R.; AVILA, L.; NEDER NETO, T.; ORFANO, B. M.. LINDSEI-BR: an oral English interlanguage *corpus*. In: **VII GSCP International Conference: Speech and Corpora**, 2013, Belo Horizonte. Proceedings of the VII GSCP International Conference: Speech and Corpora. Florença, Itália: Firenze University Press, 2012. v. 1. p. 85-86.

MENASCHE, D. **Why I'm not afraid of dying**. CNN Opinion. Disponível em: http://www.cnn.com/2014/01/08/opinion/menasche-teaching-life/index.html?iid=article_sidebar. Acesso em 10 de janeiro de 2014.

MONTRUL, S. A. The L2 acquisition of dative experiencer subjects. **Second Language Research**, 14, p. 27-61, 1998.

MURPHY, S. Second language transfer during third language acquisition. **Working Papers in TESOL & Applied Linguistics**, V. 3, N. 2, 2003.

MYLES, F. Interlanguage corpora and second language acquisition research. **Second Language Research**, 21, p. 373-391, 2005.

NUNBERG, G.; SAG, I. A.; WASOW, T. Idioms. **Language**, Volume 70, Number 3, September, p. 491-538, 1994.

NEMSER, W. Approximative systems of foreign language learners. *International Review of Applied Linguistics*, vol. 9, p.15-123, 1971. Reproduzido em Richards, J. (Ed.) **Error analysis: perspectives on second language acquisition**. Harlow: Longman Group, 1984. p. 55-63.

NESSELHAUF, N.; RÖMER, U. Lexical-grammatical patterns in *Spoken English*: the case of the progressive with future time reference. **International Journal of Corpus Linguistics**, 12(3): 297-333, 2007.

ODLIN, T. **Language transfer**: cross-linguistic influence in language learning. Cambridge: Cambridge University Press, 1989. 210p.

OH, E. **Second language acquisition of English double object constructions by korean speakers**. 2006. 330p. Tese (Doutorado em Filosofia, Linguística) - Faculty of The Graduate School, University of Southern California, California, 2006. Disponível em: <http://www-rcf.usc.edu/~ionin/SLAgroup/ohe7702578319.pdf>. Acesso em: 10 de janeiro de 2012.

OSBORNE, D. M. Systematic differences in consonant sounds between the interlanguage phonology of a Brazilian Portuguese learner of English and standard American English. **Ilha do Desterro**, Florianópolis, n. 55, p. 111-132, 2008

PALMER, G. B. **Toward a theory of cultural linguistics**. Austin: University of Texas Press, 1996.

PARK, K. Information Structure and Dative Word Order in Adult L2 Learners. In: **Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA 2011)**, ed. Julia Herschensohn and Darren Tanner, 101-109. Somerville, MA: Cascadilla Proceedings Project, 2011.

PETERNELLI, Luiz Alexandre; MELLO, Márcio Pupin de. **Conhecendo o R: uma visão estatística**. Série Didática. Minas Gerais: Editora UFV, 2011.

PIERREHUMBERT, J. B. Probabilistic phonology: discrimination and Robustness. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds). **Probabilistic linguistics**. Cambridge, Mass.: The MIT Press, 2003.

PINKER, S. **Learnability and cognition: the acquisition of argument structure**. The MIT Press, Cambridge MA, 1989.

POULISSE, N., & BONGAERTS, T. First language use in second language production. **Applied Linguistics**, 15 (1), p. 36-57, 1994.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. Disponível em: <http://www.R-project.org/>.

RStudio. 2014. Disponível em: <http://www.rstudio.com>.

RAPPAPORT HOVAV, M.; LEVIN, B. The english dative alternation: the case for verb sensitivity. **Journal of Linguistics** 44, 129-167, 2008.

RASO, T. O *corpus* C-ORAL-BRASIL. In: Tommaso RASO; Heliana MELLO. (Org.). **C-ORAL-BRASIL I. Corpus de referência do português brasileiro falado informal**. Belo Horizonte: UFMG, 2012, p. 55-90.

RASO, T.; MELLO, H.. **C-ORAL-BRASIL I**. *Corpus* de referência do português brasileiro falado informal. Belo Horizonte: UFMG, 2012. v.1, 332p

ROCHA, M. G. Síncrise pronominal: um confronto de corpora - variáveis sociolinguísticas. **Anais do II Congresso Nacional de Linguística e Filologia**, Rio de Janeiro, v. 1, p. 404-413, 1998.

SANTOS, D. Corporizando algumas questões. In: TAGNIN, S. V. O. **Avanços da linguística de corpus no Brasil**. São Paulo: Humanitas. 2008. P. 41-66.

SARDINHA, T. B. **Linguística de corpus**. São Paulo, Manole, 2004. 410 p.

SARDINHA, T.B. **Linguística de Corpus**: histórica e problemática. DELTA: Documentação de Estudos em Linguística Teórica e Aplicada. São Paulo, vol. 6, n. 2, 2000a.

SARDINHA, T. B. **O que é um corpus representativo?** DIRECT Paper Relatório Técnico 44. São Paulo / Liverpool. 2000b.

SCHACHTER, J. A new account of language transfer. In: GASS, S.; SELINKER, L.:(Eds.). **Language Transfer in Language Learning**. Amsterdam: John Benjamins Publishing, 1993. p. 32-46.

SCHER, A. P. **Construções com dois complementos no inglês e no português do Brasil**: um estudo sintático comparativo. 1996. 129p. Dissertação (Mestrado em Linguística) – UNICAMP, 1996.

SELINKER, L. **Rediscovering interlanguage**. London, New York: Longman, 1992. 288p.

_____. Interlanguage. *International Review of Applied. Linguistics*, vol. 10, p. 209-31, 1972. Reproduzido em Richards, J. C. **Error analysis**: perspectives on second language acquisition. Harlow: Longman Group, 1984.

SINCLAIR, J. **Trust the text**: language, *corpus* and discourse. London: Routledge, 2004. 224 p.

SOUSA, D. A. **Estratégias de processamento de insumo e a alternância das formas dativas em inglês**. Dissertação (Mestrado) – Instituto dos Estudos da Linguagem, Universidade Estadual de Campinas, Campinas. 2006.

SOUZA, R. A. Argument structure in L2 acquisition: language transfer re-visited in a syntax and semantics perspective. *Ilha do desterro – A journal of English language, literatures in English and cultural studies*. no. 60, 2011. p. 153-188.

Switchboard-1 Release 2. 992-2013 Linguistic Data Consortium, The Trustees of the University of Pennsylvania. Disponível em: <http://catalog.ldc.upenn.edu/LDC97S62>. Acesso em: 08 de Agosto de 2011.

TEUBERT, W. Language and *corpus* linguistics. In: TEUBERT, W.; CERMAKOVA, A. (Eds.) **Corpus linguistics: a short introduction**. London: Continuum, 2007. P. 1-58.

THEIJSEN, D. Using the ICE-GB *Corpus* to model the English dative alternation. ***Proceedings of the Aston Postgraduate Conference on Corpus Linguistics, 2008***. Disponível em: http://daphnetheijssen.ruhosting.nl/publications/Theijssen_2008_Using_ICE-GB_to_model_English_dative.pdf. Acesso em: 03 de dezembro de 2011.

THEIJSEN, D.; BRESNAN, J.; FORD, M.; BOVES, L.. **In a land far far away... A probabilistic account of the dative alternation in British, American, and Australian English**. (s/d) Sem data. Disponível em: http://daphnetheijssen.ruhosting.nl/publications/Theijssen_et_al_2011_In_a_land_far_far_away.pdf. Acesso em: 07 de julho de 2012.

TOMASELLO, M. Acquiring linguistic constructions. In: KUHN, D.; SIEGLER, R. (Eds.). **Handbook of Child Psychology**. New York: Wiley, 2006.

_____. **Constructing a language: a usage-based theory of language acquisition**. Cambridge, MA: Harvard University Press, 2003a. 388p.

_____. **Origens culturais da aquisição do conhecimento humano**. São Paulo: Martins Fontes, 2003b.

_____. The item-based nature of children's early syntactic development. **Trends in Cognitive Sciences**, vol. 4, 156-163, 2000.

VALENZUELA, J.; ROJO, A. What can foreign language learners tell us about constructions. In: DE KNOP, S.; DE RIJCKER, T. (Eds.) **Cognitive Approaches to Pedagogical Grammar** (Volume in Honour of René Dirven). Berlin-New York: Mouton de Gruyter, 2008. p.197-229

VILELA, A.C. S. Transferência linguística e transferência de treinamento na interlíngua do falante de português -L1/Inglês-L2. 2009. Dissertação (Mestrado em Estudos Linguísticos) - Faculdade de Letras da UFMG, Belo Horizonte, 2009.

WHITE, L. **Second language acquisition and universal grammar**. Cambridge: Cambridge University Press, 2003. 331p. Capítulo 7. Argument structure. p.203-239.

WHONG-BARR, M.; SCHWARTZ, B. D. Morphological and syntactic transfer in child L2 acquisition of the English dative alternation. **Studies in Second Language Acquisition**, 24, pp 579-616, 2002.

WOLK, C., WOLFER, S. A., BAUMANN, P., HEMFORTH, B., KONIECZNY, L. Acquiring English dative verbs: proficiency effects in german l2 learners. In: Carlson, L.; Hölscher, C.; Shipley, T. (Eds.). **Proceedings of the 33rd Annual Conference of the Cognitive Science Society**. Austin, TX: Cognitive Science Society, 2011.

WULFF, S.; GRIES, S. T. **Corpus-driven methods for assessing accuracy in learner production**. University of North Texas and University of California, Santa Barbara, 2006.

WURM, M.; KONIECZNY, L.; HEMFORTH, B. The acquisition of english dative alternation: proficiency effects in french L2 learners. **25th Annual CUNY Conference on Human Sentence Processing**, 2012. Disponível em: http://cuny2012.commonsgc.cuny.edu/files/2012/03/cuny2012_152.pdf. Acesso em: 04 de abril de 2012.

XIAO, R. Theory-driven *corpus* research: using corpora to inform aspect theory. In: Ludeling, A. (Ed). **Corpus linguistics: an international handbook**. Mouton de Gruyter, 2009. p. 987-1007.

ZARA, J. V. Divulgação do R para linguistas. **Anais do Encontro Virtual de Documentação em Software Livre e Congresso Internacional de Linguagem e Tecnologia Online**. v. 2, n. 1, 2013.

_____. **Estudo da expressão de eventos de transferência de posse na interlíngua de brasileiros aprendizes de inglês**. Dissertação (Mestrado). Programa de Pós-graduação em Estudos Linguísticos. Universidade Federal de Minas Gerais, 2009.

ZARA, J. V.; OLIVEIRA, F. L.; SOUZA, R. A. "Selective transfer in the acquisition of double object constructions by Brazilian learners". **Alfa: Revista de Linguística**. vol. 57, no. 2, 2013, p.

519-544.

ZEDDARI, I. **The Acquisition of the English Dative Alternation by Moroccan EFL University Learners:** an Optimality-theoretic Approach. Doctoral Dissertation, University Mohammed V, Rabat, 2009.