

Aprendizes brasileiros de inglês e sua sensibilidade aos contextos preferenciais de uso das construções dativas do inglês

Júlia Vidigal ZARA

Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas

Heliana Ribeiro de MELLO

Universidade Federal de Minas Gerais

Crysttlian Arantes PAIXÃO

Universidade Federal de Santa Catarina

Este estudo investigou a sensibilidade de aprendizes brasileiros de inglês aos seguintes fatores que influenciam a alternância dativa do inglês: pronominalidade, animacidade, acessibilidade discursiva e extensão dos argumentos recipiente e tema. Utilizando a metodologia da Linguística de Corpus (MCENERY & HARDIE 2012), nossa investigação baseou-se na análise de produções escritas e orais de falantes de inglês/L1 e da interlíngua português/ inglês extraídas de 4 corpora. Os resultados obtidos por meio dessa análise indicam que brasileiros aprendizes de inglês de níveis intermediário e avançado são capazes não somente de adquirir as construções dativas do inglês, mas também de selecioná-las em seus contextos preferenciais de uso.

Introdução

Este estudo relata os resultados de pesquisa baseada em corpora sobre a produção de construções dativas por aprendizes brasileiros de inglês de nível intermediário e avançado. As construções dativas produzidas por aprendizes de língua estrangeira têm sido objeto de estudo sob diferentes perspectivas teóricas, dadas as distintas premissas que buscam explicar os elementos linguísticos e cognitivos associados a tais construções.

Construções dativas são caracterizadas tematicamente por argumentos realizados como fonte, recipiente e tema. Tais argumentos, sintaticamente, assumem as funções de sujeito, objeto indireto e objeto direto. A estrutura construcional de um verbo dativo e seus argumentos varia tanto translinguisticamente quanto intralinguisticamente. Assim, a língua inglesa, por exemplo, exhibe tanto construções S-V-OI-OD quanto S-V-OD-OI, como ilustram as sentenças (1) e (2) abaixo:

(1) *Mark paid John a surprise visit.*

(2) *Mark paid a surprise visit to John.*

Em (1) vê-se a estrutura S-V-OI-OD, chamada de ditransitiva, enquanto em (2) observa-se a estrutura S-V-OD-OI, denominada como ditransitiva preposicionada. Além das possibilidades exibidas em (1) e (2), a língua inglesa também manifesta estruturas como em (3), frequentemente incluídas na classificação ditransitiva preposicionada, porém semanticamente claramente distintas de (1) e (2), em função da preposição utilizada:

(3) *Mark paid a surprise visit for John.*

A possibilidade de um mesmo verbo ocorrer nas construções (1) e (2) acima é chamada de "alternância dativa". Análises de corpora de inglês como língua materna (L1) revelam que a escolha por uma das construções da alternância dativa não ocorre de forma aleatória; ao contrário, é condicionada por fatores linguísticos (ex: complexidade sintática e animacidade dos complementos do verbo) e não linguísticos (ex: idade, gênero) (GRIES 2003, GRIES & STAFANOWITSCH 2004, BRESNAN ET AL. 2007, THEIJSSE 2008, BRESNAN & HAY 2008, KENDALL ET AL. 2011). Em outras palavras, estudos baseados em corpora de inglês/L1 apontam que as construções dativas do inglês, embora semanticamente semelhantes, possuem contextos preferências de uso específicos.

Poucos são os estudos, contudo, que investigam quais são os aspectos que influenciam a alternância dativa em língua inglesa-interlíngua (doravante, LII) (CALLIES & SZCZESNIAK 2008). Análises sobre a LII têm focado em tendências universais na ordem de aquisição das construções dativas do inglês (cf. CHANG 2004, HAWKINS 1987, MAZURKEWICH 1984), em efeitos de transferência de língua materna para a LII (cf. SCHWARTZ & SPROUSE 1996, ZARA 2009) e no aprendizado de quais verbos participam da alternância dativa (cf. SOUSA 2006 e ZARA 2009).

Partindo da consideração de que a proficiência em uma segunda língua (doravante, L2) envolve não apenas o conhecimento sobre suas construções e sobre as possibilidades de co-ocorrência dessas construções (ex: co-ocorrência de construções sintáticas e lexicais), mas também a capacidade de selecionar as construções da L2 em seus contextos preferenciais de uso (WULFF & GRIES 2011), buscamos neste estudo contribuir com pesquisas sobre a alternância dativa na LII investigando a sensibilidade do falante de LII brasileiro a alguns dos fatores que influenciam a preferência por uma das possibilidades construcionais da alternância dativa do inglês como língua materna. Em específico, analisamos se o falante de LII brasileiro é influenciado pelas propriedades "pronominalidade", "complexidade sintática", "animacidade" e "acessibilidade discursiva" dos argumentos recipiente (para quem algo é transferido) e tema (o que é transferido) ao escolher uma das construções dativas do inglês, de forma semelhante ao que foi demonstrado para falantes nativos de inglês (cf. estudos mencionados acima). O estudo realizado para alcançarmos nosso objetivo de pesquisa é descrito a seguir.

1. O estudo

Utilizamos neste estudo a metodologia da Linguística de Corpus (LC). A LC é uma área da Linguística que se encarrega da compilação e exploração de corpora com a finalidade de estudar a língua a partir das ocorrências de seu uso real (BIBER ET AL. 1998, MCENERY & HARDIE 2012, SARDINHA 2000). Para a LC, um corpus 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).

Nas seções abaixo, apresentamos os corpora analisados, explicando como eles foram tratados e explorados por meio de métodos estatísticos para alcançarmos nosso objetivo de pesquisa.

Conforme será exposto na discussão dos resultados alcançados e de sua interpretação, o tratamento estatístico de dados empíricos representativos é um importante aliado tanto para o estudo de fenômenos linguísticos quanto para a validação de teorias que buscam explicá-los.

1.1 Corpora analisados

Analisamos dados de inglês/L1 e LII extraídos de dois subcorpora do componente britânico do *International Corpus of English* (ICE-GB), de uma amostra do componente brasileiro do *The Louvain International Database of Spoken English Interlanguage* (LINDSEI-BR) e de uma amostra do sub-corpus brasileiro do *International Corpus of Learner English* (Br-ICLE).

O ICE-GB é um corpus representativo do inglês britânico, ideal para estudos linguísticos por ser anotado com informações sobre classes gramaticais, estrutura sintática e traços morfológicos. Compõem o corpus textos que pertencem a uma grande variedade de gêneros e registros, 200 escritos (423.702 palavras) e 300 falados (637.562 palavras). O subcorpus de textos escritos do ICE-GB será chamado neste trabalho de *Written*, e o subcorpus de textos orais de *Spoken*.

O LINDSEI—BR (MELLO et al. 2012) e o Br-ICLE¹ são corpora não-anotados da interlíngua português/inglês. Esses corpora trazem produções orais e escritas, respectivamente, de brasileiros estudantes universitários dos níveis intermediário ao avançado de proficiência em inglês/L2. No LINDSEI-BR, as produções são de fala e ocorreram durante a realização de três tipos de tarefas: uma narrativa, uma discussão livre e uma descrição. Já no Br-ICLE as produções correspondem a textos argumentativos escritos sobre tópicos variados. Ambos os corpora da interlíngua ainda estavam em fase de compilação durante a realização deste estudo, e, por essa razão, os corpora completos não foram passíveis de análise. Foi possível analisar uma amostra do LINDSEI-BR com 30.952 palavras, representando 15 falantes, e uma amostra do Br-ICLE com 159.364 palavras, distribuídas em 332 textos que correspondem a 332 aprendizes.

Os corpora supracitados foram analisados a partir da sua exploração com o auxílio do software R (R CORE TEAM 2014), apresentado na próxima seção.

1.2 Software utilizado

O R² é um software livre que tem atraído cada vez mais pesquisadores pelas vantagens que oferece (PETERNELLI & MELLO 2011). Por exemplo, o R permite, em um só ambiente, a manipulação de dados, a realização de cálculos e a geração de gráficos (GRIES 2009). Além disso, o pesquisador utilizando o R tem mais controle sobre a análise que está sendo feita, podendo instruir o R através de scripts. É possível, por exemplo, definir como uma busca por informações específicas será realizada em um arquivo de texto com características particulares.

Criamos scripts do R neste trabalho para manipular os arquivos de corpus (ex: para eliminar informações irrelevantes para esta pesquisa), contar o número de palavras desses arquivos, realizar análises estatísticas e gerar gráficos. Além disso, o R facilitou a localização das construções dativas, que foram extraídas dos corpora segundo os critérios especificados a seguir.

1.3 Critérios para a seleção de expressões dos corpora

A seleção de expressões dos corpora foi feita com base em critérios semânticos e sintáticos. Primeiramente, foram localizados nos corpora, com a ajuda do R, os verbos que poderiam expressar transferência literal ou metafórica de pose e que interagem tanto com a construção ditransitiva quanto com a ditransitiva preposicionada do inglês (ex: *give*, *tell*). Em seguida, foi observado, através da análise dos contextos de ocorrência dos verbos, se os verbos de fato ocorriam nessas construções. Verbos interagindo com outras construções, não relevantes para este estudo, foram eliminados, como ilustram os exemplos abaixo:

.
(4) “*We question whether Bishop Lowth is telling the truth*”. → verbo + tema

(5) “*I told you that I was required (...)*” → verbo + recipiente + that-clause

(6) “*(...) what I was told by several (...)*” → construção passiva

(7) “*Her husband told her not to attend (...)*” → verbo + recipiente + infinitive

(8) “*(...) tell me if it is (...)*” → verbo + recipiente + if-clause

.

Foram excluídas também sentenças do *subcorpus* oral de inglês/L1 (*Spoken*) com recipientes e/ou temas que foram marcados como não compreendidos durante a transcrição do *corpus* (9):

.

(9) “*I haven’t given them <unclear-words> <unclear-words>*”

.

No que tange à construção ditransitiva preposicionada em específico, foram selecionadas ocorrências nas quais o referente do sintagma preposicional era semanticamente compatível com o papel de <recipiente>, literal ou metafórico, ou de <beneficiário>, em oposição a uma leitura apenas locativa desse constituinte (10-11):

.

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

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

.

As sentenças selecionadas com base nos critérios acima foram categorizadas conforme explicamos a seguir.

1.4 Categorias de análise das sentenças selecionadas

Do ponto de vista sintático, cada sentença foi marcada como contendo a construção ditransitiva ou a ditransitiva preposicionada. Além disso, os argumentos <recipiente> e <temas> de cada construção dativa selecionada foram classificados quanto às propriedades abaixo, definidas conforme BRESNAN & HAY (2008):

I. Pronominalidade: refere-se à condição de um argumento ser expresso através de um sintagma cujo núcleo é um pronome pessoal (12), um pronome demonstrativo (13) ou um pronome indefinido (14), como ilustram os seguintes exemplos do subcorpus *Written* do ICE-GB:

.

(12) “*I ‘ve told you the truth.*”

(13) *"Why have I ended up quoting you this?"*

(14) *"Isn't it about time you started telling everyone the truth?"*

.

II. Complexidade sintática: corresponde ao número de palavras. No exemplo (15) abaixo, extraído do *Written*, considera-se que o argumento <tema> (5 palavras) é mais extenso do que o argumento <recipiente> (1 palavra):

.

(15) *"(...) 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.

.

(16) *"(...) giving a young person <animado> the opportunity to watch those in work <inanimado>" (Written).*

.

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.

.

(17) *"(...) is to grant these early texts <given> secondary status <new> (...)"*

.

A partir das categorias definidas para a análise das ocorrências foi possível realizar as análises quantitativas que passamos a descrever na próxima seção.

1.5 Análise quantitativa dos dados

Os primeiros dados quantitativos calculados dizem respeito às frequências de ocorrência total de cada construção dativa por corpus. Nosso objetivo foi observar se os universos analisados eram semelhantes ou diferentes quanto à distribuição das construções foco do nosso estudo. Em seguida, definimos as distribuições de cada tipo de recipiente e de tema por corpus, buscando relacionar essas distribuições às frequências das construções.

Comparamos as frequências obtidas por meio de testes de significância estatística a fim de verificarmos se as diferenças observadas entre as frequências eram fruto do acaso ou não, isto é, se elas apresentavam significância estatística ou não. Esses testes geram um valor-p que é comparado com o nível de significância (α) estabelecido para o teste. Quando o valor-p é menor que o nível de significância, a diferença entre as frequências observadas é considerada estatisticamente

significativa.

O nível de significância geralmente utilizado em testes estatísticos é igual a 0,05. Em casos de múltiplas comparações, contudo, é necessário utilizar um nível de significância corrigido ($\alpha\beta$), por meio do procedimento de Bonferroni. A correção de Bonferroni é realizada com a utilização da fórmula $\alpha\beta = \alpha / k$, sendo $\alpha\beta$ o nível de significância ajustado, α o nível de significância original e k o número de comparações realizadas (CORDER & FOREMAN 2009:81). Nos resultados apresentados neste estudo, especificamos abaixo de cada tabela o nível de significância utilizado.

Foi necessário recorrermos a dois testes de significância estatística diferentes para analisarmos conjuntos de dados com características distintas. Um deles foi o teste qui-quadrado, recomendado quando todas as frequências observadas são maiores que 1 e quando pelo menos 80% das frequências são maiores que 4 (GRIES 2013:166). O outro foi o Teste Exato de Fisher, mais preciso do que o qui-quadrado quando muitas frequências são pequenas (MCDONALD 2009).

O próximo passo da análise quantitativa foi calcular, por *corpus*, as frequências de ocorrência de cada tipo de recipiente e de tema em cada construção dativa. Com isso, buscamos identificar, em cada *corpus*, tendências de associações entre tipos específicos de recipientes e temas com as construções dativas, e comparar, entre os *corpora*, se as associações observadas eram semelhantes ou não.

Paratestarmosestatisticamenteseos *corpora* analisados se aproximavam ou se distanciavam uns 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, fizemos uma análise de *cluster*. Esse tipo de análise é utilizado para se classificar um conjunto de elementos (ex: 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).

Após a análise de cluster, no intuito de explorarmos de forma mais detalhada a alternância dativa na LII brasileira, apresentamos também resultados sobre a associação entre cada propriedade do recipiente (animacidade, pronominalidade, acessibilidade e extensão) e as construções dativas do inglês separadamente. Focamos no argumento <recipiente> baseados na conclusão de Gries (2003), que destaca as propriedades desse argumento como as que possuem maior poder discriminatório entre as duas construções dativas do inglês. Segundo o autor, na construção ditransitiva prototípica, o argumento <recipiente> é curto, pronominal, animado e dado no discurso.

Vejam na próxima seção os resultados alcançados por meio da metodologia apresentada acima.

2. Resultados e discussão

Os números totais de expressões selecionadas do *Spoken*, do *Written*, do Br-ICLE e do LINDSEI-BR foram 516, 340, 191 e 21, respectivamente. Não surpreendentemente, quanto maior o *corpus*, maior o número de ocorrências obtidas. Resultados provenientes de universos de tamanhos tão diferentes podem ser comparados graças a testes estatísticos que comparam os conjuntos de dados proporcionalmente (MCENERY & WILSON, 2001:85).

A primeira característica que observamos nos universos analisados foi a predominância da construção ditransitiva. Como se pode observar na Tabela 1, a construção ditransitiva representa 79,71% das sentenças analisadas do *Written*, 82,56% das sentenças extraídas do *Spoken* e 62,83% do total de sentenças selecionadas no Br-ICLE. A quarta coluna da Tabela 1 mostra os resultados da aplicação do teste qui-quadrado, utilizado para comparar as frequências observadas em cada

corpus.

	Ditransitiva	Ditransitiva Preposicionada	Valor-p obtido (qui-quadrado)
<i>Written</i>	79,71% (271)	20,29% (69)	< 0,0001 (≠)
<i>Spoken</i>	82,56% (426)	17,44% (90)	< 0,0001 (≠)
Br-ICLE	62,83% (120)	37,17% (71)	< 0,0001 (≠)
LINDSEI-BR	61,90% (13)	38,10% (8)	0,2170 (=)

Table 1. TABELA 1: Distribuição das construções dativas por corpus e valores-p obtidos por meio do teste qui-quadrado para a comparação entre as frequências. Destaque para valores-p que indicam diferenças estatisticamente significativas entre as frequências observadas. $\alpha = 0,05$

Os resultados apresentados na Tabela 1 revelam, portanto, a primeira semelhança entre os dados de inglês/L1 e LII: o corpus escrito da interlíngua (Br-ILCE) apresenta a mesma tendência dos corpora de inglês/ L1 (*Written* e *Spoken*) no que diz respeito à frequência estatisticamente maior da ditransitiva em relação à ditransitiva preposicionada. Embora o valor-p obtido para o LINDSEI-BR (valor-p = 0,217) aponte para a não relevância estatística da diferença observada entre a frequência da ditransitiva (61,90%) e da ditransitiva preposicionada (38,10%) nesse corpus, acreditamos que um resultado semelhante ao obtido para o *Spoken* teria sido alcançado se tivéssemos analisado uma amostra maior do *corpus* oral de interlíngua.

A grande ocorrência da ditransitiva apontada acima pode ser explicada se atentarmos para os tipos de recipientes e temas que ocorreram com mais frequência nos universos que analisamos, pois eles são exatamente aqueles apontados por GRIES (2003) como favorecendo a seleção da construção ditransitiva: recipientes pronomes, animados, dados discursivamente e menores que os temas, e temas NPs lexicais, inanimados e expressando informações novas.

Vejamos as distribuições dos diferentes tipos de recipiente e tema nos corpora considerando cada propriedade separadamente. Começando pela pronominalidade, recipientes pronomes foram estatisticamente mais frequentes que recipientes NP lexicais no *Written*, no *Spoken* e no LINDSEI-BR. 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. Temas, por sua vez, foram estatisticamente mais frequentes como NPs lexicais do que como pronomes (Tabela 3). Os valores-p seguidos do símbolo ≠ nas Tabelas 2 e 3 indicam diferenças estatisticamente significativas entre as frequências comparadas.

	Recipiente pronome	Recipiente MP lexical	Recipiente pronome	Recipiente NP lexical	Valor-p
<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%	0,3083 (≠)
LINDSEI-BR	16	5	76,19%	23,81%	0,0029 (≠)

Table 2. TABELA 2: Distribuição de recipientes “pronomes” e “NP lexicais” nos corpora de inglês/L1 e L2 e valores-p resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências observadas. Destaque para frequências estatisticamente maiores. $\alpha = 0,05$

	Tema pronome	Tema MP Lexical	Valor-p
<i>Written</i>	5,29% (18)		0,0004 (≠)
<i>Spoken</i>	11,24% (58)		0,0004 (≠)
Br-ICLE	4,19% (8)		0,0004 (≠)
LINDSEI-BR	19,05% (4)		0,0001 (≠)

Table 3. TABELA 3: Distribuição de temas “pronomes” e “NPs lexicais” nos corpora de inglês/L1 e L2 e valores-p 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 observadas. Destaque para frequências estatisticamente maiores. $\alpha = 0,05$

Passando para a propriedade “animacidade”, em todos os corpora, houve maior ocorrência total de

recipientes animados e temas inanimados (Tabela 4). As diferenças primeiramente observadas entre as frequências relativas à propriedade “animacidade” foram bastante evidentes e posteriormente confirmadas por meio do teste qui-quadrado e do teste exato de Fisher. 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, com valor-p resultante da comparação entre essas duas frequências igual a 0,0004 ($\alpha = 0,05$). Similarmente, das 191 ocorrências de recipientes no Br-ICLE, 82,40% são animados (ex: “*Should we give them food*”) e 17,60% são inanimados, com valor-p também igual a 0,0004.

	Recipiente animado	Recipiente inanimado	Tema animado	Tema inanimado
Written	86,80% (294)	13,20% (46)	0,59% (2)	99,41% (338)
Spoken	95,16% (491)	4,84% (25)	17,44% (90)	82,56% (426)
Br-ICLE	82,72% (158)	17,28% (33)	1,05% (2)	98,95% (189)
LINDSEI-BR	100,00% (21)	0,00% (0)	0,00% (0)	100,00% (21)

Table 4. TABELA 4: Distribuição de recipientes e temas animados e inanimados nos corpora

Outra característica comum a todos os corpora foi a predominância de recipientes “dados” (Tabela 5), ou seja, recipientes que correspondem a informações já mencionadas anteriormente, como em “*I’m not even bothering to send them postcards*”, extraído do *Written*. Nos corpora escritos de inglês/L1 e da interlíngua, por exemplo, 82,94% e 82,72% dos recipientes são dados, respectivamente. Os temas, por sua vez, tenderam a ser informações novas com significância estatística no *Spoken*, no *Written* e no Br-ICLE.

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

Table 5. TABELA 5: Distribuição de recipientes dados e novos em cada corpus de inglês/L1 e L2 e valores-p resultantes da aplicação do teste qui-quadrado para a comparação entre as frequências observadas. Destaque para frequências estatisticamente maiores. ($\alpha = 0,05$)

A análise da propriedade “extensão” também aponta para um fator comum entre os corpora *Written*, *Spoken* e Br-ICLE que favorece a seleção da ditransitiva: os recipientes foram predominantemente menores que os temas em 72,94%, 75,78% e 63,35% das vezes, respectivamente (Tabela 6). Essa tendência não foi observada apenas no LINDSEI-BR, no qual o número de recipientes menos extensos que temas, 33,33%, não foi estatisticamente maior do que o número de recipientes mais extensos que temas, 9,52%. A não significância estatística entre essas duas frequências foi revelada pelo valor-p igual a 0,1300 gerado por meio da aplicação do Teste Exato de Fisher (Tabela 7).

	recipiente < tema	recipiente > tema	recipiente = tema
Written	72,94% (248)	12,35% (42)	14,71% (50)
Spoken	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)

Table 6. TABELA 6: 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.

	Written	Spoken	Br-ICLE	LINDSEI-BR
Recipiente < tema versus Recipiente > tema	< 0,0001 (≠)	< 0,0001 (≠)	< 0,0001 (≠)	0,1300 (=)
Recipiente < tema versus Recipiente=tema	< 0,0001 (≠)	< 0,0001 (≠)	< 0,0001 (≠)	0,2146 (=)
Recipiente > tema versus Recipiente=tema	0,4325 (=)	< 0,0001 (≠)	0,69153 (=)	0,0025 (≠)

Table 7. TABELA 7: Valores-p 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 os valores-p que indicam não haver diferença estatisticamente significativa entre as frequências comparadas. $\alpha\beta = 0,0166$

Mostramos acima as distribuições dos diferentes tipos de recipientes no universo de sentenças consideradas neste estudo. Argumentamos, com base em GRIES (2003), que essas distribuições explicam a maior ocorrência da construção ditransitiva nos *corpora* que analisamos, já que recipientes foram predominantemente pronomes, animados e dados, e temas foram predominantemente inanimados, novos e mais extensos. Esses resultados já apontam para a sensibilidade dos aprendizes a alguns dos fatores que favorecem a seleção da construção ditransitiva quando a ditransitiva preposicionada seria gramaticalmente possível.

Vejamos agora os resultados que revelam de forma mais direta as associações observadas nos *corpora* entre as propriedades dos recipientes e temas consideradas em nossa análise e as construções dativas do inglês. Primeiramente, o resultado da análise de *cluster* (Figura 1) mostra como os *corpora* aproximam-se ou distanciam-se uns dos outros em relação aos tipos de recipientes e temas mais frequentemente identificados em cada construção dativa.

No lado superior direito do *heatmap*³ (Figura 1) visualizamos a maior proximidade entre a ditransitiva preposicionada no *Spoken*, no *Written* e no Br-ICLE. Isso significa que foram identificados nesses três *corpora* os mesmos tipos de recipientes e temas ocorrendo na construção ditransitiva preposicionada.

Portanto, podemos dizer que os falantes representados no *Spoken*, no *Written* e no Br-ICLE demonstraram uma preferência semelhante pela escolha da construção ditransitiva preposicionada quando os mesmos tipos de recipientes e temas foram utilizados. Focando agora no lado esquerdo do *heatmap*, observamos a maior proximidade entre ocorrências da construção ditransitiva no *Spoken*, no *Written* e no Br-ICLE no que diz respeito aos tipos de recipientes e temas observados nas sentenças analisadas.

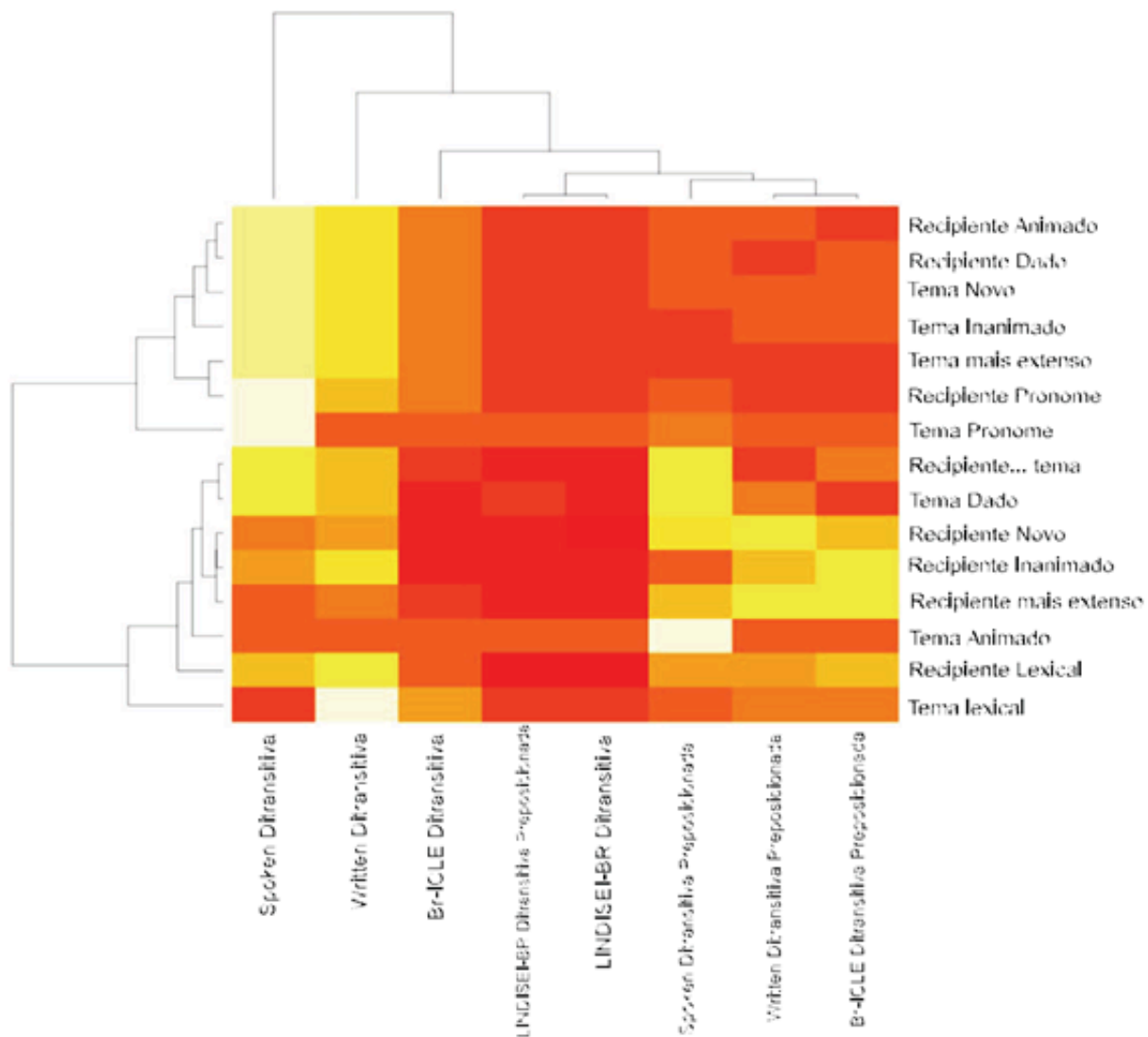


Figure 1. FIGURA 1: Heatmap gerado como resultado da análise de cluster para comparar as distribuições dos diferentes tipos de recipientes e de temas que ocorrem nas construções ditransitiva e ditransitiva preposicionada nos corpora de inglês L1 e L2

Vimos que o aspecto que predominantemente determinou os agrupamentos do *heatmap* 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 os falantes nativos do inglês bem como os aprendizes representados nos *corpora* por nós analisados diferenciam de forma semelhante os usos das construções dativas do inglês sob a influência das propriedades animacidade, pronominalidade, acessibilidade e extensão dos argumentos recipientes e temas.

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 o menor número de dados do *corpus* oral foi o responsável por esse resultado.

Na análise de *cluster* discutida acima consideramos as propriedades “pronominalidade”, “animacidade”, “acessibilidade” e “extensão” dos recipientes e temas simultaneamente. Passemos agora para os resultados de testes de significância estatística que olham para a associação entre

cada propriedade dos recipientes e temas e as construções dativas na LII separadamente.

Em relação à propriedade “pronominalidade”, observamos nos *corpora* da LII a mesma associação entre recipientes pronomes e a construção ditransitiva identificada nos *corpora* de inglês/L1. Como mostra a Tabela 8, a construção ditransitiva foi a escolhida em 86,14%, 81,25%, 92,27% e 89,46% das vezes em que um recipiente foi expresso através de um pronome no Br-ICLE, no LINDSEI-BR, no *Written* e no *Spoken*, respectivamente.

A aplicação do teste exato de Fisher mostrou que não há diferença estatisticamente significativa entre os *corpora* de inglês L1 e L2 quanto à frequência de recipientes pronomes na construção ditransitiva. Na Tabela 9 podemos visualizar os valores-p maiores que o nível de significância 0,0083 que mostram que tal diferença não existe, indicando que os aprendizes representados nesta pesquisa são tão sensíveis quanto os falantes nativos da língua-alvo à associação entre recipientes pronomes e a construção ditransitiva.

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

Table 8. TABELA 8: Distribuição de recipientes pronomes por construção 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 (=)

Table 9. TABELA 9: 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 $\alpha = 0,0083$

Diferenças entre os *corpora* foram identificadas, porém, quando olhamos para os resultados relativos aos recipientes NPs lexicais, tanto quando comparamos os *corpora* de inglês L1 entre si, quanto quando comparamos os *corpora* de inglês L1 com os *corpora* da interlíngua (Tabela 10). Se, por um lado, a construção ditransitiva foi a mais frequente quando o recipiente foi expresso por um NP lexical no *Written* (valor-p = 0,0039), por outro lado, não houve diferença estatisticamente significativa entre a frequência de recipientes NPs lexicais na ditransitiva e na ditransitiva preposicionada no *Spoken* (valor-p = 0,0829). Uma tendência diferente foi ainda encontrada nos *corpora* de aprendizes, com uma associação maior 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. No LINDSEI-BR, essa escolha ocorreu em 100% das vezes. É possível, então, levantar a hipótese de que os aprendizes representados nesta pesquisa tendem a associar recipientes expressos através de NPs lexicais mais à construção ditransitiva preposicionada do que à construção ditransitiva.

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)
valor-p	0,0039 (\neq)	0,0829 (=)	0,0009 (\neq)	0,0079 (\neq)

Table 10. TABELA 10: Frequências de recipientes NP lexicais por construção em cada corpus de inglês L1 e L2 e os valores-p resultantes da aplicação do teste qui-quadrado (*Written*, *Spoken* e Br-ICLE) e do teste exato de Fisher (LINDSEI-BR) para a comparação entre essas frequências. Destaque para frequências estatisticamente maiores $\alpha = 0,05$

No que tange à propriedade “animacidade”, encontramos a mesma tendência nos *corpora* de inglês L1 e L2 em relação à associação entre recipientes animados e a construção ditransitiva. Na Tabela 11 podemos ver que a proporção de recipientes animados que ocorreram na construção ditransitiva foi de 82,99%, 83,30%, 74,66% e 61,90% no *Written*, no *Spoken*, no Br-ICLE e no LINDSEI-BR,

respectivamente.

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

Table 11. TABELA 11: Distribuição de recipientes animados por construção nos corpora de inglês/L1 e L2

Ao compararmos os corpora entre si no que diz respeito às frequências observadas na Tabela 11, não constatamos diferenças estatísticas significativas entre eles, como mostram os valores-p da Tabela 12. Interpretamos esses resultados como evidências de que os aprendizes, assim como os falantes nativos de inglês, são influenciados pela animacidade do recipiente ao escolherem uma das construções dativas do inglês.

	Spoken	Br-ICLE	LINDSEI-BR
Written	0,9270 (=)	0,0429 (=)	0,0234 (=)
Spoken		0,0164 (=)	0,0234 (=)
Br-ICLE			0,3208 (=)

Table 12. TABELA 12: Valores-p 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. Valores-p maiores que 0,0083 indicam que $\alpha\beta = 0,0083$

Outra semelhança entre os corpora de inglês L1 e L2 foi constatada quando olhamos para a propriedade “acessibilidade discursiva” do recipiente. Como mostra a Tabela 13, 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, em todos os corpora.

Written	Recipiente dado	Recipiente novo
Ditransitiva	88,30% (249)	37,93% (22)
Ditr. prep.	11,70% (33)	62,07% (36)
valor-p	0,0004 ≠	0,0209 ≠
Spoken	Recipiente dado	Recipiente novo
Ditransitiva	87,15% (407)	38,78% (19)
Ditr. prep.	12,85% (60)	61,22% (30)
valor-p	0,0004 ≠	0,0404 ≠
LINDSEI-BR	Recipiente dado	Recipiente novo
Ditransitiva	81,25% (13)	0,00% (0)
Ditr. prep.	18,75% (3)	100,00% (5)
valor-p	0,0010 ≠	0,0079 ≠
Br-ICLE	Recipiente dado	Recipiente novo
Ditransitiva	73,42% (116)	12,12% (4)
Ditr. prep.	26,58% (42)	87,88% (29)
valor-p	0,0004 ≠	0,0004 ≠

Table 13. TABELA 13 - Frequências de recipientes animados e inanimados por construção e valores-p resultantes da comparação entre essas frequências por meio dos testes qui-quadrado e exato de Fisher*. Destaque para frequências estatisticamente maiores $\alpha = 0,05$

Observamos, contudo, uma diferença entre o Br-ICLE e os corpora de inglês L1 quando comparamos os corpora entre si quanto à distribuição de recipientes dados na ditransitiva e na ditransitiva preposicionada (Tabela 14). Uma vez constatada diferença estatisticamente significativa entre as frequências comparadas, podemos dizer que, embora os aprendizes sejam influenciados pela acessibilidade discursiva do recipiente ao selecionarem uma das construções

dativas do inglês, a associação entre recipientes dados e a construção ditransitiva é menor no *corpus* escrito de aprendizes do que nos *corpora* de falantes nativos do inglês que analisamos.

	Written	Br-ICLE	LINDSEI-BR
Spoken	0,7317 (=)	0,0001 (≠)	0,4507 (=)
Written		0,0001 (≠)	0,4223 (=)
Br-ICLE			0,7647 (=)

Table 14. TABELA 14: Valores-p 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. Valores-p menores que 0,0083 indicam diferenças entre os *corpora* comparados $\alpha = 0,0083$

Por fim, consideremos os resultados obtidos para o fator “extensão”. Estudos anteriores mostram que recipientes menores que temas favorecem o uso da construção ditransitiva (GRIES, 2003). Corroborando esses resultados anteriores, os números exibidos na Tabela 15 revelam que, em todos os *corpora* por nós analisados, recipientes menores que temas favorecem o uso da construção ditransitiva. Essa associação ocorreu em 85,12% das vezes no Br-ICLE (ex: “give us reasons for living and fighting for our plans”) e 85,71% das vezes no LINDSEI-BR (ex: “offer me like a lot of courses that I could take”), por exemplo.

Comparamos os 4 *corpora* entre si com relação às distribuições de recipientes menores que temas por construção (Tabela 16). Encontramos semelhanças entre os *corpora* escritos da L1 e da interlíngua (valor-p = 0,1509), e também entre os *corpora* orais da L1 e da interlíngua (valor-p = 0,3068). Por outro lado, identificamos diferenças entre os *corpora* oral e escrito de inglês L1 (valor-p = 0,0329) e entre o *corpus* oral da L1 e o escrito da interlíngua (*Spoken* e Br-ICLE). É possível, portanto, que haja uma influência do canal na força da associação entre recipientes mais curtos e a construção ditransitiva. O fato de não terem sido observadas diferenças entre o LINDSEI-BR e os *corpora* escritos pode ser consequência do tamanho da amostra do *corpus* oral da interlíngua.

		recipiente < tema
Spoken	Ditransitivas	94,88% (371)
	Ditr.preposicionadas	5,12% (20)
	Valor-p	0,0004 (≠)
Written	Ditransitivas	90,32% (224)
	Ditr.preposicionadas	9,68% (24)
	Valor-p	0,0004 (≠)
Br-ICLE	Ditransitivas	85,12% (103)
	Ditr.preposicionadas	14,88% (18)
	Valor-p	< 0,0004 (≠)
LINDSEI-BR	Ditransitivas	85,71% (6)
	Ditr.preposicionadas	14,29% (1)
	Valor-p	0,02914 (≠)

Table 15. Tabela 15: Frequências de recipientes menores que temas por construção nos *corpora* de inglês L1 e L2 e valores-p resultantes da aplicação dos testes qui-quadrado e exato de Fisher para a comparação entre essas frequências. Destaque para frequências estatisticamente maiores $\alpha = 0,05$

	Written	Br-ICLE	LINDSEI-BR
Spoken	0,0329 (=)	0,0009 (≠)	0,3068 (=)
Written		0,1509 (≠)	1,0000 (=)
Br-ICLE			1,0000 (=)

Table 16. TABELA 16: Valores-p resultantes da aplicação do teste exato de Fisher para a comparação entre os *corpora* de inglês L1 e L2 quanto às distribuições de recipientes menores que temas por construção. Destaque para valores-p que levam à rejeição da hipótese nula $\alpha = 0,05$

Conclusão

Conforme demonstrado através da análise estatística dos dados dos *corpora* aqui estudados, há uma tendência semelhante entre os *corpora* de inglês/L1 e da interlíngua português/inglês no que diz respeito à diferenciação que os falantes fazem entre os contextos preferenciais de uso da construção ditransitiva e da ditransitiva preposicionada do inglês quando consideramos as propriedades “pronominalidade”, “complexidade sintática”, “animacidade” e “acessibilidade discursiva” dos argumentos <recipiente> e <tema>. Portanto, nossos resultados indicam que aprendizes de L2 podem alcançar um conhecimento da língua-alvo que vai além do aprendizado das formas das construções nela existentes, incluindo uma sensibilidade aos contextos preferenciais de uso dessas construções. Destacamos que esse tipo de conhecimento é necessário para atingir níveis de proficiência mais avançados na L2 (WULFF & GRIES, 2011).

Além disso, é importante notar que, ao demonstrarem sensibilidade aos contextos preferenciais de uso das construções dativas do inglês, os aprendizes representados nos *corpora* que analisamos foram capazes de capturar informações probabilísticas do insumo linguístico (co- ocorrência entre tipos específicos de argumentos e construções sintáticas) que provavelmente não foram apresentadas a eles por meio de instrução formal, o que pode ser interpretado como evidência favorável às teorias linguísticas probabilísticas (cf. BOD, HAY & JANNEDY 2003). Ressaltamos que a análise de dados empíricos quantitativamente relevantes deve ser a base para a confirmação ou rejeição de teorias linguísticas.

É necessário também um comentário final sobre algumas diferenças observadas entre os *corpora* de inglês L1 e L2, como, por exemplo, a maior associação entre recipientes NP lexicais e a construção ditransitiva preposicionada no *corpus* escrito da interlíngua (Br-ICLE) do que no *corpus* escrito de inglês/L1 (*Written*). Diferenças como essas entre os *corpora* de falantes nativos e de aprendizes de inglês não podem ser vistas como inesperadas se considerarmos resultados de estudos que comparam *corpora* de falantes nativos da L1 entre si. Por exemplo, BRESNAN & HAY (2008), KENDALL et al. (2011) e THEIJSEN et al. (s/d) mostraram que a comparação entre *corpora* de falantes nativos de inglês que pertencem a grupos diferentes em termos de idade, gênero e variedade do inglês revela diferenças entre os grupos de falantes no que diz respeito ao grau de importância das variáveis que influenciam a seleção por uma das construções dativas do inglês. Portanto, considerando que os falantes/escritores representados nos *corpora* de inglês L1 e L2 analisados neste estudo pertencem a grupos diferentes em diversos aspectos, podemos dizer que é esperado encontramos alguma diferença entre os *corpora* analisados em termos do grau de associação entre algumas propriedades dos argumentos <recipiente> e <tema> e as construções sintáticas analisadas. Afinal, a linguagem humana é suficientemente complexa e constantemente influenciada por diversos fatores para deixar de apresentar tal variabilidade. Cabe a nós, linguistas, desenvolvermos métodos de análise para capturar e explicar tal variabilidade.

References

1. BIBER Douglas, CONRAD Susan, REPPEN Randi. *Corpus linguistics: investigating language structure and use*. Cambridge University Press: Cambridge; 1998.
2. BOD Rens, HAY Jennifer, JANNEDY Stefanie. *Probabilistic linguistics*. The MIT Press: Cambridge, Mass; 2003.
3. BRESNAN Joan , et al. Predicting the Dative Alternation. In: Bouma Gerlof, Kraemer Irene , Zwarts Joost . *Cognitive foundations of interpretation*. Royal Netherlands Academy of Science: Royal Netherlands Academy of Science; 2007.
4. BRESNAN Joan, Hay Jennifer. Gradient Grammar: an effect of animacy on the syntax of give in New Zealand and American English. *Lingua*. 2008; 118:245-259.
5. CALLIES Marcus , SZECZESNIAK Konrad. Argument realisation, information status and syntactic weight - A learner-corpus study of the dative alternation. In: WALTE Maik, GROMMES Patrick. *Fortgeschrittene Lernervarietäten Korpuslinguistik und Zweitsprachenerwerbsforschung*. Walter de Gruyter: Walter de Gruyter; 2008:165-187.

6. CANTOS GOMEZ Pascual. Do we need statistics when we have linguistics?. *D.E.L.T.A.* 2002; 18(2):233-271.
7. CHANG Lan-Hsin. Discourse effects on EFL learners' production of dative constructions. *Journal of National Kaohsiung University of Applied Sciences.* 2004; 33:145-170.
8. CORDER Gregory , FOREMAN Dale. *Nonparametric Statistics for Non-Statisticians: a Step-by-Step Approach.* Wiley: New Jersey; 2009.
9. FEW Stephen. *Multivariate Analysis Using Heatmaps. Visual Business Intelligence Newsletter.* 2006.
10. GRIES Stefan Th. *Statistics for linguistics with R.* De Gruyter Mouton: Berlin & New York; 2013.
11. GRIES Stefan Th. *Quantitative corpus linguistics with R: a practical introduction.* Routledge: New York; 2009.
12. GRIES Stefan Th. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics.* 2003; 1:1-27.
13. GRIES Stefan Th, Stefanowitsch Anatol. Extending collocation analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics.* 2004; 9(1):97-129.
14. HAWKINS Roger. Markedness and the acquisition of the English dative alternation by L2 speakers. *Second Language Research.* 1987; 3:21-55.
15. KENDALL Tyler, BRESNAN Joan, HERK Gerard van. The dative alternation in African American English: Researching Syntactic Variation and Change across Sociolinguistic Datasets. *Corpus linguistics and linguistic theory.* 2011; 7(2):229-244.
16. MAZURKEWICH Irene. The acquisition of the dative alternation by second language learners and linguistic theory. *Language Learning.* 1984; 34:91-109.
17. MCDONALD John. *Handbook of biological statistics.* Sparky House Publishing: Baltimore, Maryland; 2009.
18. MCENERY Tony, HARDIE Andrew A.. *Corpus linguistics: method, theory and practice.* Cambridge University Press: Cambridge; 2012.
19. MCENERY Tony, WILSON Andrew. *Corpus linguistics: an introduction.* Edinburgh University Press: Edinburgo; 2001.
20. MELLO Heliana , et al. LINDSEI-BR: an oral English interlanguage corpus. In: VII GSCP International Conference: Speech and Corpora, 2012, Belo Horizonte. *Proceedings of the VII GSCP International Conference: Speech and Corpora.* Firenze University Press: Firenze University Press; 2012:85-86.
21. PETERNELLI Luiz Alexandre, MELLO Márcio Pupin de. *Conhecendo o R: uma visão estatística.* Editora UFV: Minas Gerais; 2011.
22. R CORE TEAM. *R: A language and environment for statistical computing.* 2014.
23. SARDINHA Tony Berber. *Linguística de corpus.* Manole: São Paulo; 2004.
24. SARDINHA Tony Berber. *O que é um corpus representativo?.* 2000.
25. SCHWARTZ Bonnie, SPROUSE Rex. L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research.* 1996; 12:40-72.
26. SOUSA Denilson Amade. *Estratégias de processamento de insumo e a alternância das formas dativas em inglês.* 2006.
27. THEIJSSSEN Daphne. Using the ICE-GB Corpus to model the English dative alternation. *Proceedings of the Aston Postgraduate Conference on Corpus Linguistics.* 03 de julho de 2014.
28. THEIJSSSEN Daphne, et al. In a land far far away... A probabilistic account of the dative alternation in British, American, and Australian English. 07 de julho de 2014.
29. WULFF Stefanie, GRIES Sefan Th. Corpus-driven methods for assessing accuracy in learner production. In: ROBINSON Peter. *Second language task complexity: researching the Cognition Hypothesis of language learning and performance.* John Benjamins: John Benjamins; 2011.
30. ZARA Júlia Vidigal. *Divulgação do R para linguistas. Anais do Encontro Virtual de Documentação em Software Livre e Congresso Internacional de Linguagem e Tecnologia Online.* 2013.



31. ZARA Júlia Vidigal. Estudo da expressão de eventos de transferência de posse na interlíngua de brasileiros aprendizes de inglês. 2009.