

As construções com *SE* na produção escrita de brasileiros aprendizes de espanhol como língua estrangeira: um estudo baseado em *corpus*¹

Constructions with *SE* in the written production of Brazilian learners of Spanish as a foreign language: a corpus-based study

Benivaldo José de Araújo Júnior*

RESUMO: Este trabalho tem por objetivo analisar a produção de construções com o clítico *SE* (doravante construções-*SE*) em língua espanhola por aprendizes brasileiros de Espanhol como língua estrangeira (ELE). Nessa análise, cujo referencial teórico é a Gramática Cognitiva, serão tratadas em especial as construções reflexivas, médias, impessoais e passivas. Após breve introdução teórica, apresentamos nosso *corpus* de estudo, exibimos e discutimos os resultados do levantamento das construções-*SE* nesse *corpus*. Na discussão, comparamos os dados que obtivemos com aqueles observados em outros dois *corpora* de falantes nativos, um para o Espanhol na variedade peninsular e outro para o Português Brasileiro (doravante PB). Finalmente, com base nessa comparação, fazemos algumas considerações sobre a produção das construções-*SE* em nosso *corpus* de estudo e os fatores que possivelmente têm influência nesse processo.

PALAVRAS-CHAVE: Construções-*SE*. ELE. Gramática Cognitiva. Estudos comparados. Linguística de *Corpus*.

ABSTRACT: This work aims to analyze the production of *SE*-constructions in Spanish by Brazilian learners of Spanish as a Foreign Language (SFL). In this analysis, which uses Cognitive Grammar as theoretical framework, reflexive, middle, impersonal and passive constructions will be specially examined. After a brief theoretical introduction, we present our corpus of study; then, we demonstrate and discuss the results of *SE*-constructions survey in this corpus. In the discussion, our data are compared with those observed in two native speakers' corpora, one for Iberian Spanish and other for Brazilian Portuguese (hereinafter PB). Finally, based on that comparison, we make some considerations on the production of *SE*-constructions in our corpus of study and the factors that possibly influence this process.

KEYWORDS: *SE*-constructions. SFL. Cognitive Grammar. Comparative studies. Corpus Linguistics.

1. O modelo de conceitualização adotado

Na classificação e interpretação dos nossos dados, adotaremos o modelo cognitivo do evento canônico (LANGACKER, 1991, p. 285). Este, por sua vez, é a combinação de dois outros: o modelo da bola de bilhar e o modelo do palco. De acordo com o primeiro (bola de bilhar), tendemos a conceitualizar os eventos do mundo como uma cadeia de ação, na qual um

*Doutor em Língua Espanhola pela FFLCH-USP e professor da Escola Superior de Propaganda e Marketing (ESPM-SP).

¹ Este trabalho é uma versão ampliada e modificada de outro apresentado no II Congresso Internacional de Professores de Línguas Oficiais do MERCOSUL (Buenos Aires, 2013), intitulado “As construções com *SE* na produção escrita de brasileiros aprendizes de espanhol como língua estrangeira”.

elemento energético transfere energia e provoca efeitos em outro elemento tipicamente não energético. Quanto ao segundo (palco), relaciona-se à experiência perceptual e foi assim chamado porque nele o papel do conceitualizador é similar ao de um espectador assistindo a uma peça; ou seja, alguém que se encontra fora de cena, mas que constitui parte do evento global. No modelo do evento canônico (fig. 1), segundo a terminologia de Langacker (1991, p. 283), o iniciador da cadeia de ação (fonte de energia) chama-se trajetor e é também o sujeito/agente prototípico; o participante final (ralo por onde escoia a energia) chama-se marco e é também o objeto/paciente prototípico; o conceitualizador/observador está indicado na figura pela letra O. Conforme o modelo, no enunciado *Ernesto rompió la estatuilla* [PB: Ernesto quebrou a estatueta], *Ernesto* é o iniciador (fonte)/trajetor/sujeito/agente e *estatuilla* é o participante afetado (ralo)/marco/objeto/paciente.

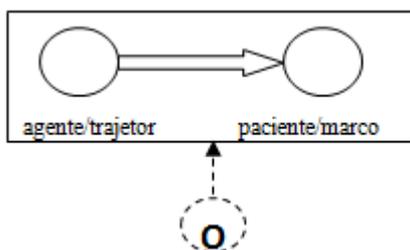


Fig.1 – Modelo de evento canônico
(LANGACKER, 1991, p. 285)

O modelo de evento canônico representa a observação usual de uma ação prototípica nas línguas do sistema nominativo-acusativo, como o PB e o espanhol, sendo, portanto, o modelo mais produtivo para nossa análise. Portanto, no modelo que adotamos, um evento canônico é transitivo por definição e nele temos dois participantes claramente distintos numa relação assimétrica, no qual o participante agente é animado e volitivo e afeta direta e totalmente o participante paciente por meio de uma força ou transferência de energia (GIVÓN, 1984, apud KEMMER, 1994).

2. As construções reflexivas e médias segundo Kemmer (1994)

Conforme já apontado por Hopper & Thompson (1980, p. 277), nos eventos em cuja codificação morfossintática aparece o pronome *SE/SE*², a assimetria entre os participantes é problemática, acarretando uma baixa na transitividade. Por exemplo, na sentença *Clara se vio*

² A notação *SE* (itálico) refere-se à ocorrência de construções com clítico em espanhol ou na produção não nativa nessa língua. A notação *SE* (normal) se aplica à incidência de construções com clítico no PB.

en el espejo (PB: *Clara se viu no espelho*), os papéis de trajetor/agente e marco/paciente são preenchidos pela mesma entidade referencial, de modo que a distinguibilidade entre esses dois participantes esmaece. Contudo, para Kemmer (1994, p. 207), é possível operar uma separação conceitual entre *Clara* como iniciadora da atividade (o participante que vê) e seu reflexo no espelho (o que é visto), de modo a poder distinguir dois participantes no evento. O mesmo ocorre com o enunciado *La presidenta se imaginó descansando en una playa desierta* [PB: *A presidenta se imaginou descansando numa praia deserta*]. Neste caso, o sujeito interage com uma representação de si mesmo situada em algum espaço mental distinto do discurso (MALDONADO, 2006, p. 270). A esse tipo de eventos, Kemmer (1994, p. 207) classifica como **reflexivos**. Ainda segundo a autora (1994), nos eventos reflexivos o iniciador (trajetor) atua sobre si mesmo como o faria sobre qualquer outra entidade.

Há ainda eventos nos quais a distinção conceitual entre os participantes é mínima. No exemplo *Felipe solo se afeita los domingos* [PB: *Felipe só se barbeia/faz a barba aos domingos*], o paciente em questão não é unicamente afetado: barbear o próprio rosto implica a atividade da parte diretamente envolvida (a cabeça), além de outras partes do corpo (o pescoço). Em eventos dessa natureza, ambos os participantes têm como referente uma única entidade cujos aspectos são conceitualmente indistinguíveis. Kemmer (1994) classifica esses eventos como **médios**, fundamentando-se no conceito de **voz média** (originalmente empregado para designar uma categoria inflexional das línguas clássicas indoeuropeias), cuja função, em termos semânticos, é “expressar eventos nos quais a ação ou estado afeta o sujeito do verbo ou seus interesses” (LYONS, 1970, p. 286). Após minucioso estudo translíngüístico, Kemmer enumera dez situações que configuram um domínio semântico no qual a voz média apresenta algum tipo de marcação morfológica. Por motivos de concisão, daremos apenas exemplos do espanhol e do PB, nos quais o *SE/SE* funciona como marcador médio: (1) cuidados corporais (*peinarse; frotarse* [PB: *pentear-se; esfregar-se*]); (2) movimento não translacional (*inclinarse* [PB: *inclinarse*]); (3) mudança na postura corporal (*levantarse; sentarse* [PB: *levantar-se; sentar-se*]); (4) movimento translacional (*irse* [PB: *ir-se*]); (5) eventos naturalmente recíprocos (*abrazarse; besarse* [PB: *abraçar-se; beijar-se*]); (6) média indireta (*preguntarse* [PB: *perguntar-se*], referente a uma atividade mental); (7) média de emoção (*enfadarse* [PB: *irritar-se*]); (8) discurso emotivo (*quejarse* [PB: *queixar-se*]); (9) média de cognição (*acordarse* [PB: *lembrar-se*]); (10) eventos espontâneos (*originarse* [PB: *originar-se*]). No PB, em muitas dessas situações, coexistem as formas com realização do

clítico ou seu apagamento, como nos enunciados *Meu pai (se) levanta cedo* e *Os fieis (se) sentaram*.

Kemmer (1994) esquematiza os eventos reflexivos e médios conforme as figuras 2 e 3.



Fig. 2 – Esquema do evento reflexivo
(KEMMER, 1994, p. 207)

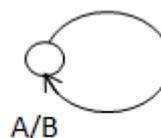


Fig. 3 – Esquema do evento médio
(KEMMER, 1994, p. 207)

Em sua análise sobre transitividade, Hopper & Thompson (1980, p. 277) propõem as construções reflexivas como intermediárias entre as transitivas e as intransitivas. Kemmer (1994, p. 209) amplia essa proposta, incluindo no conjunto as construções médias e utilizando o grau de distinguibilidade dos participantes como parâmetro semântico para diferenciar uma construção da outra. Dessa forma, a autora constrói um diagrama no qual num dos extremos estão os eventos intransitivos (1 participante), e no outro os transitivos (2 participantes). Entre esses dois extremos, estão os eventos nos quais a distinção entre os participantes é baixa, caso dos reflexivos e médios.

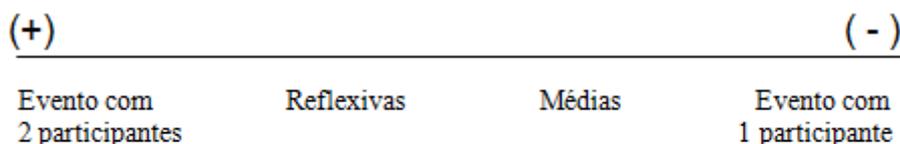


Fig. 4 – Grau de distinguibilidade dos participantes (KEMMER, 1993, p. 73)

Segundo o diagrama, nas construções reflexivas/ recíprocas, pelo fato de conceitualmente podermos distinguir dois participantes (embora se refiram à mesma entidade), temos maior proximidade a um evento transitivo. Já nas construções médias, a baixa distinguibilidade entre os participantes aproxima-as das intransitivas. Kemmer (1994) relaciona a maior ou menor distinguibilidade dos participantes com uma propriedade semântica a qual denomina **elaboração relativa de eventos**.

A marcação morfológica das reflexivas/recíprocas e médias no PB e no espanhol, quando ocorre, normalmente se dá com o mesmo marcador *SE/SE*, conforme já mostrado em

exemplos anteriores. Nas referidas línguas, portanto, a distinção entre essas construções não é imediata e requer uma análise acurada do verbo em termos semânticos. No caso do PB, salientamos a existência de um fator que complica ainda mais essa classificação: a perda dos clíticos (inclusive o SE) que vem acontecendo nessa língua³.

3. As construções passivas e impessoais segundo Maldonado (2006)

Para Maldonado (2006, p. 273), que se concentra fundamentalmente na análise dessa questão na língua espanhola, a presença do *SE* nas construções médias, passivas e impessoais se relaciona com a força indutiva envolvida no evento.

Com base nessa hipótese, o autor propõe a seguinte escala de classificação (figura 5): num dos extremos da escala estariam as construções transitivas, com máxima proeminência de transferência de energia; no outro extremo, estariam as construções absolutas, sem energia envolvida; na zona intermediária, as médias, passivas e impessoais⁴.

Nessa classificação, além dos graus de manifestação da força indutiva, são levados em conta o requisito de agentividade humana do verbo, o aspecto (léxico e morfológico), a concordância e a ordem. Com base nesses critérios, Maldonado (2006) classifica como médias algumas construções consideradas passivas por muitos gramáticos e linguistas, como é o caso de D; para esse autor, tal passividade é aparente, uma vez que resulta mais da perda de proeminência da força indutora que da escolha do tema como figura do evento.

³ Esse fenômeno foi analisado em diversos estudos. González (1994) leva em conta a perda dos clíticos no PB, ao falar das assimetrias entre o espanhol e o PB na descrição do preenchimento vs. não preenchimento das posições argumentais de sujeito e objeto nessas línguas. Não trataremos dos casos de perda, visto que o foco do nosso trabalho são as construções nas quais aparece o *SE*; porém, colocaremos o clítico entre parênteses toda vez que a construção puder aparecer sem ele.

⁴ As impessoais intransitivas (*Se vive bien aquí* [PB: *Vive-se bem aqui*]) não apresentam ambiguidades quanto à classificação no E e no PB; portanto, neste trabalho, receberão mais atenção as impessoais com objeto direto (*Se alquila(n) casas* [PB: *Aluga(m)-se casas*]).

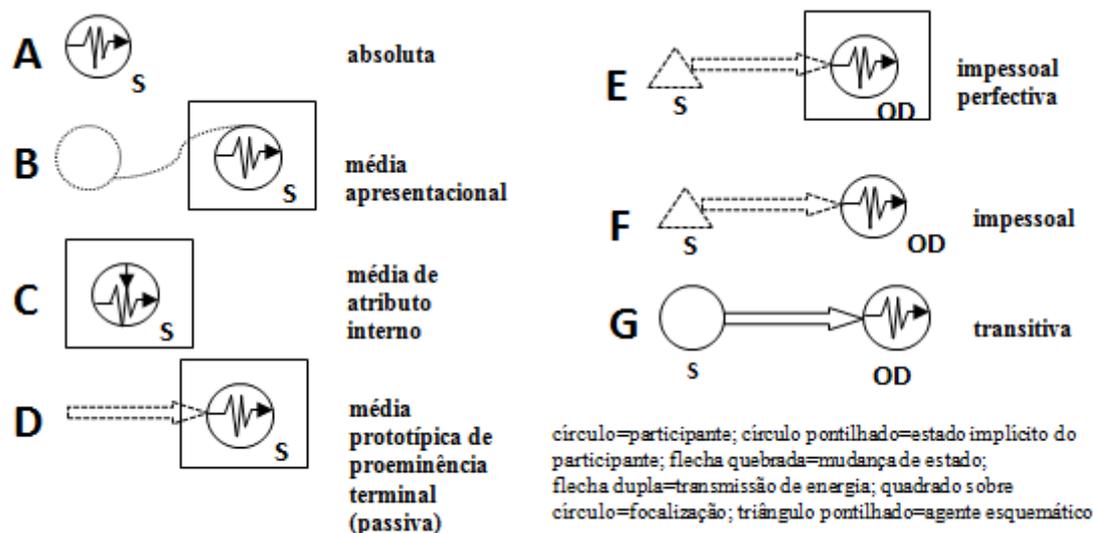


Fig. 5 – Níveis de proeminência da força indutiva (MALDONADO, 2006, p. 273)

A seguir, definimos sucintamente as construções da figura 5:

- (a) **Absoluta:** sem energia em perfil. Ex.: *El ruido disminuyó* [PB: *O ruído diminuiu*].
- (b) **Média apresentacional:** nível quase nulo de energia; simplesmente mostra um evento num domínio qualquer a partir do ponto de vista de um conceitualizador. Ex.: *Las Islas Fiji se encuentran en el Pacífico* [PB: *As Ilhas Fiji se encontram no Pacífico*].
- (c) **Média de atributo interno:** a mudança de estado sofrida pelo tema depende mais de sua configuração interna que de uma força externa. Ex.: *Estos juegos no se venden bien* [PB: *Estes jogos não (se) vendem bem*].
- (d) **Média prototípica de proeminência terminal (Passiva):** focaliza a porção terminal do evento (a mudança de estado devido a uma força externa esquemática) e não sua fase indutiva. Em nossa análise, utilizaremos a forma abreviada MPT para referir-nos a essa construção; igualmente, manteremos entre parênteses a denominação **passiva**, por ser a mais frequente na literatura. Ex.: *La pared se manchó* [PB: *A parede (se) manchou*].
- (e) **Impessoal perfectiva:** a indução da força é tão importante quanto a mudança de estado, e tem que ser humana, embora seu agente não seja específico. Ex.: *Se pagó la deuda* [PB: *Pagou-se a dívida/ A dívida foi paga*⁵].

⁵ Conforme mostrado em Araújo Júnior (2006), os enunciados com SE ocorrem no PB, porém a preferência dos falantes nativos dessa língua é pela passiva perifrástica.

(f) **Impessoal:** os marcadores imperfectivos reduzem a proeminência da mudança de estado e favorecem o que de homogêneo possa haver no evento; portanto, o que se perfila são as tendências naturais de mudança sofridas pelo tema. Ex.: *Estos productos se fabrican con materiales sintéticos* [PB: *Estes produtos se fabricam/ são fabricados com materiais sintéticos*].

(g) **Transitiva:** prototipicamente, são as construções nas quais um agente animado e volitivo transfere energia a um paciente, provocando neste uma mudança de estado. Ex.: *Miguel pintó la casa* [PB: *Miguel pintou a casa*].

4. Os corpora

Nosso *corpus* de estudo é o Dados EEC (doravante DEEC), que possui 178.066 palavras e está constituído por 1.172 produções escritas de alunos do curso *Español en el Campus*⁶. O DEEC foi construído para integrar um *corpus* maior, o COMET (*Corpus Multilíngue para Ensino e Tradução*), desenvolvido pelo CITRAT⁷ com o objetivo de servir de suporte a pesquisas linguísticas, principalmente nas áreas de tradução, terminologia e ensino de línguas. O objetivo específico do DEEC era coletar e organizar dados longitudinais, que possibilitassem acompanhar um grupo de informantes do primeiro ao último estágio do EEC. A coleta de produções ocorreu entre os meses de agosto de 2003 e junho de 2006, e abarcou os níveis Básico, Intermediário e Avançado. Na tabela a seguir, mostramos uma síntese do perfil dos informantes do DEEC.

⁶ Curso de ELE oferecido pela Área de Língua Espanhola e Literaturas Espanhola e Hispano-americana – DLM e mantido pelo Serviço de Cultura e Extensão da FFLCH/USP entre 1996 e 2010. A grade completa compreendia os níveis Básico I (B1), Básico II (B2), Intermediário I (I1), Intermediário II (I2), Avançado I (A1) e Avançado II (A2).

⁷ Centro Interdepartamental de Tradução e Terminologia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

Tabela 1 – DEEC: Perfil dos informantes.

		QUANTIDADE	%
INFORMANTES POR SEXO	Masculino	50	27
	Feminino	136	73
	TOTAL	186	100
INFORMANTES POR FAIXA ETÁRIA	Até 20 anos	22	11,8
	De 21 a 34 anos	98	52,7
	De 35 a 49 anos	45	24,2
	De 50 a 69 anos	21	11,3
	TOTAL	186	100

Quanto à escolaridade, 94,6% do total de informantes chegou à universidade. No geral, esses colaboradores pertenciam à comunidade USP: eram estudantes (graduandos ou pós-graduandos), professores ou funcionários da instituição. Do total, apenas 4% afirmavam ter tido algum tipo de contato com a língua espanhola antes de ingressar no EEC.

As produções do *corpus* pertencem a diversos gêneros discursivos⁸, tais como anúncios pessoais, cartas formais e informais, diários, resenhas, crônicas, artigos de opinião, notícias, sinopses, críticas de filmes, etc. Na tabela a seguir, disponibilizamos a quantidade de produções do DEEC segundo o nível dos informantes.

Tabela 2 – DEEC: Produções por nível.

NÍVEL	QUANTIDADE	%	Nº DE PALAVRAS
B1	545	46,5	62.376
B2	251	21,4	47.627
I1	147	12,5	22.514
I2	121	10,3	20.268
A1	60	5,1	14.420
A2	48	4,1	10.861
TOTAL	1.172	100	178.066

⁸ Convém esclarecer que os gêneros presentes no *corpus* não são autênticos, mas sim projeções: não se trata, por exemplo, de jornalistas redatando crônicas, críticas ou notícias para um público leitor; o que se tem é o exercício de escrita em língua estrangeira (no caso, em espanhol) feito por alunos, mediante instrução formal e tendo como único destinatário o professor. Portanto, embora se tenha buscado a diversidade nas práticas textuais pedidas aos alunos, buscando aproximá-las dos gêneros correntes no cotidiano, essas produções são inevitavelmente atravessadas pela prática pedagógica.

Os resultados da análise das construções-*SE* no DEEC serão cotejados com aqueles observados em outros dois *corpora* de falantes nativos, um para o espanhol e o outro para o PB. Ambos são *corpora* já constituídos e na modalidade oral. No caso do espanhol, trata-se do *PRESEEA-ALCALÁ*, que por sua vez é parte integrante do *PRESEEA* (*Proyecto para el Estudio Sociolingüístico del Español de España y de América*)⁹. As 36 entrevistas que compõem esse *corpus* foram realizadas em 1998, na cidade de Alcalá de Henares¹⁰. Na tabela 3 aparecem informações detalhadas sobre o *corpus* *PRESEEA-ALCALÁ*, ao qual nos referiremos doravante como CPA.

Tabela 3 – *Corpus* PRESEEA-ALCALÁ: Resumo das principais características.

Corpus PRESEEA-ALCALÁ (CPA)			
Faixa etária	Sexo	Nível de escolaridade	
		Superior (Nº de	Médio (Nº de
Faixa 1 20-34 anos	masculino	24.730	24.257
	feminino	24.442	26.471
	Total da faixa 1	49.172	50.728
Faixa 2 35-54 anos	masculino	30.932	26.311
	feminino	22.882	25.443
	Total da faixa 2	53.814	51.754
Faixa 3 Acima de 55 anos	masculino	29.695	29.860
	feminino	33.940	30.398
	Total da faixa 3	63.635	60.258
Total por nível de escolaridade		166.621	162.740
Total geral		329.361	

Para o PB, utilizamos a Amostra SP2010 (Piloto), que é parte do *corpus* que está sendo desenvolvido pelo GESOL-USP (Grupo de Estudos e Pesquisa em Sociolinguística da USP) com o objetivo de subsidiar pesquisas na área de sociolinguística variacionista — em especial, na descrição e análise dos fenômenos variáveis e processos de mudança em curso na variedade paulistana do PB (MENDES, 2011). As 36 entrevistas que constituem esse *corpus*

⁹ Projeto surgido em 1993 e que conta com equipes de pesquisa e documentação na Espanha, nos Estados Unidos e em alguns países americanos (Argentina, Colômbia, Cuba, Chile, Equador, Guatemala, México, Paraguai, Peru, Porto Rico, Uruguai e Venezuela). Fonte: <http://linguas.net/portalpreseea/Inicio/tabid/441/language/es-ES/Default.aspx>, consultado em 20/05/13.

¹⁰ A informação completa acerca da metodologia de coleta desses materiais está em Moreno Fernández et alii (2002, 2004).

foram realizadas na cidade de São Paulo, entre 2008 e 2011. A tabela 4 mostra em detalhe as principais características da Amostra SP2010, à qual nos referiremos doravante como ASP¹¹.

Tabela 4 – Amostra SP2010 (Piloto) (GESOL-USP): Resumo das principais características.

Amostra SP2010 (Piloto)(GESOL-USP)			
Faixa etária	Sexo	Nível de escolaridade	
		Superior (Nº de	Médio (Nº de
Faixa 1 20-34 anos	masculino	28.145	30.414
	feminino	35.411	37.170
	Total da faixa 1	63.556	67.584
Faixa 2 35-49 anos	masculino	36.330	33.977
	feminino	32.627	29.416
	Total da faixa 2	68.957	63.393
Faixa 3 50-69 anos	masculino	31.884	32.730
	feminino	27.737	33.357
	Total da faixa 3	59.621	66.087
Total por nível de escolaridade		192.134	197.064
Total geral		389.198	

Escolhemos o CPA e a ASP para este estudo, primeiramente, porque foram constituídos a partir dos mesmos critérios metodológicos. Em segundo lugar, ambos os *corpora* são recentes e, portanto, refletem o estado atual do espanhol e do PB com maior probabilidade de incorporar fenômenos emergentes nessas línguas. Por último, os dois *corpora* apresentam pouca disparidade no quesito extensão: a ASP (389.198 palavras) supera o CPA (329.361 palavras) em apenas 18%; ambos são *corpora* médios, conforme a classificação proposta por Berber Sardinha (2004, p. 26)¹².

Embora as tabelas 3 e 4 ofereçam dados acerca da faixa etária, sexo e escolaridade dos informantes, tais variáveis não serão levadas em conta em nossa análise. O que aqui faremos é uma primeira aproximação a alguns aspectos referentes à produção de construções-SE em língua espanhola por aprendizes brasileiros de ELE; neste nível, portanto, nos limitaremos a analisar a incidência global e os percentuais dessas construções no *corpus* de estudo e compará-los com os resultados obtidos no CPA e na ASP.

¹¹ As tabelas 3 e 4 são uma versão simplificada daquelas presentes em Araújo Júnior (2013, p. 131-2).

¹² Segundo esse autor, estão nesse grupo os *corpora* que possuem entre 250.000 e 1 milhão de palavras.

5. Metodologia e análise do *corpus* de estudo

O levantamento das construções-*SE* no *corpus* DEEC foi feito em três etapas, conforme o procedimento que se descreve a seguir. A etapa inicial constou de três passos, sendo o primeiro dividir o *corpus* em seis *subcorpora* — B1, B2, I1, I2, A1, A2 —, considerando os níveis da grade (ver tabela 2). A seguir, processamos cada *subcorpus* no programa Kitconc¹³, obtendo, assim, as concordâncias por meio da lematização **SE*. O Kitconc está escrito totalmente em português e reúne algumas funcionalidades usadas em Linguística de *Corpus*, tais como a listagem de palavras, frequências, concordâncias, colocados e a extração de palavras-chave; no caso, usamos o programa apenas como concordanciador. A lematização **SE* nos permitiu acessar todas as combinações do clítico no *corpus*, tanto as proclíticas (*SE levanta*, *SE hacía*, *SE ha comprado*, etc.) quanto as enclíticas (*marcharSE*, *verSE*, *realizándoSE*, etc.). O terceiro passo foi transferir as concordâncias para uma planilha Excel, a fim de facilitar sua leitura e triagem.

A segunda etapa, executada manualmente, consistiu em ler as concordâncias e eliminar aquelas sem interesse para o trabalho. Portanto, foram desprezadas as ocorrências do *SE* como variante de *LE/LES* (objeto direto em construções do tipo *SE la dio*, *SE lo entregó*, etc.) e outros casos (*faSE*, *eSE*, *informaSE*, etc.).

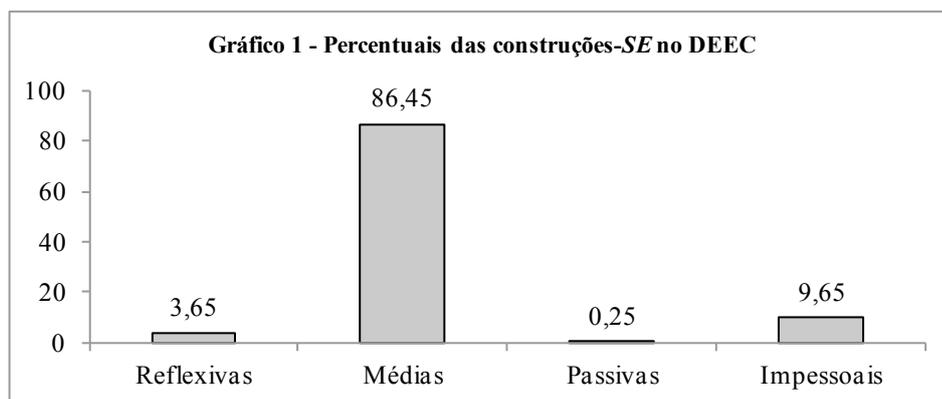
A etapa final consistiu em classificar as incidências do *SE* nas concordâncias remanescentes. Embasando-nos em Kemmer (1994) e Maldonado (2006), reunimos essas ocorrências em reflexivas, médias, passivas (MPT) e impessoais. O resultado final está na tabela a seguir:

Tab. 5 – DEEC: Totais por construção.

Subcorpus	Reflexivas	Médias	Passivas (MPT)	Impessoais	TOTAL
B1	6	644	0	21	671
B2	39	371	0	20	430
I1	1	158	0	21	180
I2	6	95	0	49	150
A1	5	68	1	26	100
A2	2	61	3	19	85
TOTAL	59	1.397	5	155	1.616
%	3,65	86,45	0,31	9,59	100

¹³ Desenvolvido por José Lopes Moreira Filho e disponível como *freeware* na Internet. Neste trabalho, utilizamos a versão 3.0 do programa.

Observando os resultados de modo global, é notória a supremacia das médias (86,45%) com relação às demais construções-SE no *corpus*. Em segundo lugar vêm as impessoais (9,65%), seguidas das reflexivas (3,65%) e, finalmente as passivas (0,25%). No gráfico a seguir, esses percentuais podem ser melhor visualizados:



A alta incidência das médias pode ser explicada em função dos critérios de classificação adotados. Ou seja, delimitamos as construções reflexivas e médias a partir da proposta de Kemmer (1994), fundamentada na elaboração relativa de eventos e no grau de distinguibilidade de seus participantes (ou dos subeventos componentes); dessa forma, em nossa análise, classificamos como médias algumas construções que normalmente são consideradas reflexivas (sobretudo nas abordagens prescritivas) ou recebem outra denominação. A seguir, citamos exemplos do *corpus*¹⁴ para algumas das situações mediais listadas por Kemmer (1994):

- (1) *Hay que dejarlos se ducharen todos los días* (ED400494) [cuidados corporais]
- (2) *Entonces se levantó en un rato, paró un poco porque estaba fatigado* (EC300013) [mudança na postura corporal]
- (3) *A los 23 años, él se fué a vivir en Salvador* (EC300353) [movimento translacional]
- (4) *En ese momento, ella se alarmó: “Pero, ¿estaré más tiempo sola, sin un novio?”* (EC300071) [média de emoção]

As reflexivas totalizaram pouco menos de 4% do total. Considerando-se apenas o conjunto dessas construções no *corpus*, houve incidência de reflexivas recíprocas (maioria,

¹⁴ A informação entre parênteses, após cada exemplo, representa o código da produção escrita no DEEC.

com 83%), reflexivas diretas (15%) e reflexivas indiretas (2%), representadas, respectivamente, nos exemplos (5), (6) e (7).

(5) *Se quieren y se respetan* (ED101572)

(6) *A ella le gusta olharse en el espejo* (EA100384)

(7) (...) *se están preparando unos pasteles para comérselos* (EC300074)

No levantamento das impessoais e passivas, utilizamos a proposta de Maldonado (2006), embasada na força indutiva presente nos eventos. Embora fatores como a concordância e o aspecto sejam importantes na diferenciação dessas construções, o requisito de agentividade humana é preponderante: se a indução no evento é humana, embora esquemática, há maior probabilidade de termos uma construção impessoal (esquemas E e F na fig. 5); se a força indutiva tem menos proeminência no evento, a saliência passa a ser do tema e estamos no âmbito das médias e passivas (esquemas B, C e D na fig. 5).

As impessoais totalizaram cerca de 10% do total de incidências. Levando-se em conta apenas o conjunto dessas construções, obteve-se no *corpus* a seguinte distribuição: as impessoais intransitivas correspondem a 22% (exemplo 8); as impessoais com complemento direto, que foram maioria, totalizam 78%. Destas, 75% são imperfectivas (exemplo 9) e 3% são perfectivas (exemplo 10), segundo o aspecto verbal.

(8) (...) *se produce más y mejor en casa* (...) (ED400271)

(9) (...) *sólo se requiere buena capacidad de comunicación* (ED400171)

(10) (...) *es una cosa que nunca se ofreció* (EF600024)

As passivas constituem apenas 0,25% do total. Para essas construções, citamos os exemplos a seguir.

(11) (...) *no la compre [la mermelada] si ya se rompió el lacre* (EE500021)

(12) (...) *se extraviaron mis maletas desde hace una semana* (EF600021)

Nessas ocorrências, destacamos a pouca saliência da força indutiva, cuja fonte específica não é identificável: a ruptura do lacre ou o extravio das maletas podem não ter sido provocados por agentes humanos; os agentes — humanos ou não — são inespecíficos na codificação do evento, de modo que o elemento proeminente em cada construção é o tema (*el*

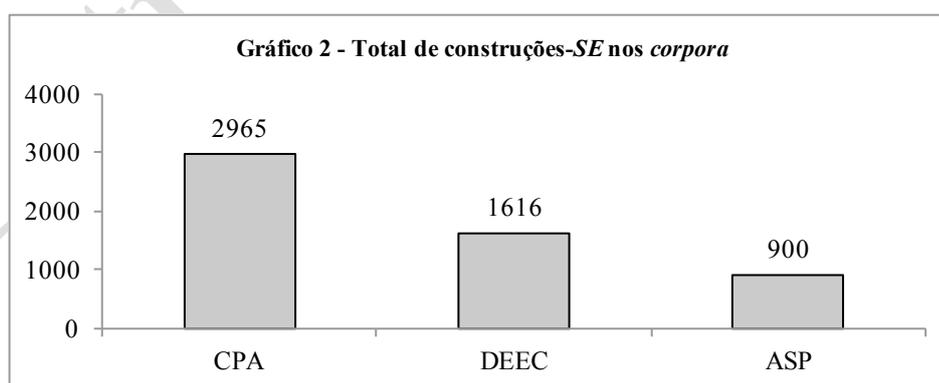
lacre, las maletas). Enfim, o que se focaliza nas passivas é a mudança de estado sofrida pelo tema e não o processo em si; daí o aspecto perfectivo presente nessas construções.

Cotejando os dados do DEEC com os do CPA (falantes nativos do espanhol peninsular) e da ASP (falantes nativos do PB), chegamos à seguinte tabela:

Tab. 6 – CPA/ DEEC/ ASP: Totais por construção.

CORPORA		Reflexivas	Médias	Passivas (MPT)	Impessoais	TOTAL
CPA	total	98	1.651	12	1.204	2.965
	%	3,31	55,68	0,4	40,61	100
DEEC	total	59	1.397	5	155	1.616
	%	3,65	86,45	0,31	9,59	100
ASP	total	96	626	2	176	900
	%	10,67	69,56	0,22	19,56	100

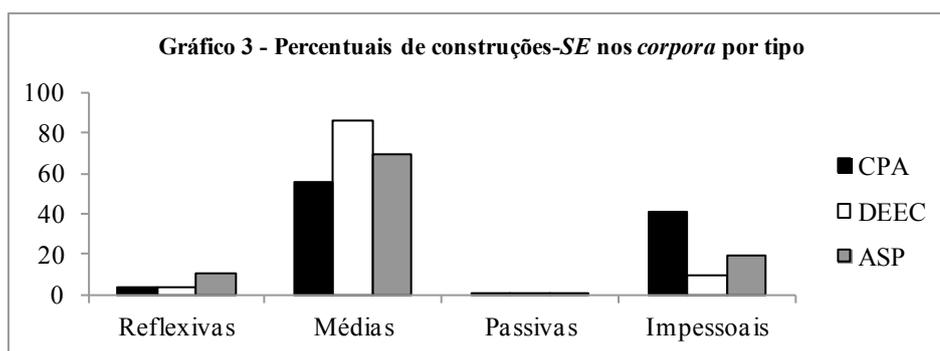
Comparando-se o total das construções-SE nos 3 *corpora*, o maior número de incidências ocorreu no CPA: o total nesse *corpus* supera em mais de 3 vezes o observado na ASP e, em quase o dobro, o total levantado no DEEC (gráfico 2). Tal resultado está dentro do esperado para a ASP, conforme os estudos comparativos de Araújo Júnior (2006) e González (1994), segundo os quais a perda ou apagamento dos clíticos (SE e os demais) no PB favoreceria a menor incidência de construções-SE nessa língua e na produção não nativa de lusoparlantes em espanhol.



Os resultados no DEEC, tomados globalmente, estão a meio caminho entre o CPA e a ASP: se por um lado a instrução formal em LE influenciaria a presença significativa de construções-SE no *corpus* de aprendizes (quase o dobro da que aparece no *corpus* de falantes

nativos do PB), por outro lado, o apagamento do clítico implicaria uma frequência ainda aquém da observada no *corpus* do espanhol.

Ao comparar os percentuais por construção (gráfico 3), os resultados se mostram difusos — o CPA lidera na quantificação percentual das impessoais e passivas; a ASP tem a maior porcentagem de reflexivas e o DEEC está à frente no percentual das médias — e não sinalizam propriamente uma tendência.



Se a perda do clítico pode estar ocasionando os baixos percentuais das passivas no DEEC¹⁵ (0,31%) e na ASP (0,22%), nas médias parece ocorrer o contrário: a alta incidência dessas construções no DEEC (86,45%) e na ASP (69,56%) sinaliza menor índice de apagamento do clítico na variedade em questão.

Sobre outros fenômenos observados no *corpus*, destacamos o baixo índice de construções sem concordância (0,43%). Citamos duas delas na sequência:

(13) *Existe el sitio donde se fabrica las piezas* (ED400701)

(14) (...) *habia formaciones minerales como aquellas que se ve en cavernas de la Tierra* (EB200302)

No caso, nossa expectativa era de que o enfraquecimento da concordância no PB (GALVES, 1993, apud GONZÁLEZ, 1994) resultasse na produção de mais construções discordantes no *corpus* de aprendizes, porém tal não ocorreu; não se pode atribuir o fato à influência da instrução formal em LE unicamente, uma vez que no próprio *corpus* do PB esse índice também foi baixo (0,8%). Outro fenômeno verificado no *corpus* de aprendizes,

¹⁵ Em Araújo Júnior (2006), ao analisarmos o *corpus* DEEC, constatamos que a perda do clítico também favoreceria a predominância das passivas perifrásticas nesse *corpus*.

também apontado por González (1994), foi a supergeneralização dos clíticos (exemplos 15 e 16), que consiste na presença destes onde não deveriam aparecer. No caso específico do *SE*, computamos 2,1% dessas construções no *corpus*.

(15) *Pero no fue eso que se sucedió* (EF600744)

(16) *Para el bebé, todo no se pasa de un juguete* (EC300283)

Conforme González (1994, p. 411), tais formas são incorporadas à produção não nativa muito mais pelo efeito sonoro do que propriamente pela importância na construção do sentido e na referencialidade. Também observamos ocorrências no DEEC nas quais o *SE* não desempenha função argumental (exemplos 17 e 18), mas cuja presença pode ser creditada à influência da língua materna, uma vez que o fenômeno ocorre com relativa frequência no PB.

(17) *Ésta es una pregunta difícil de se contestar* (EF501114)

(18) *Sólo había piedras y más piedras para se estudiar* (EC201152)

Nesses casos — que totalizaram 0,8% no DEEC e 2,3% na ASP — o *SE/SE* acompanha infinitivos em construções impessoais e parece ter como finalidade reforçar o caráter humano do sujeito esquemático na estrutura.

5. Considerações finais

A análise do *corpus* de aprendizes que aqui procedemos, embora sucinta, permite-nos tecer algumas considerações.

De início, ressaltamos que a Gramática Cognitiva foi uma abordagem produtiva na análise do nosso *corpus* de estudo, uma vez que nos ofereceu critérios (especialmente semânticos) que nos auxiliaram a melhor delimitar e classificar as construções-*SE* levantadas. De acordo com os dados, a incidência global das construções-*SE* no *corpus* de aprendizes ainda está distante do observado no espanhol, possivelmente influenciada pela perda do clítico no PB. Entretanto, a análise das ocorrências por construção revelou baixo índice de passivas e alto percentual das médias no *corpus* de aprendizes; consideramos esses resultados um indício de que o apagamento do clítico no PB ocorreria de maneira diferenciada nas diferentes construções-*SE*, e que uma investigação específica desse fenômeno ofereceria mais respostas sobre a produção das referidas construções, em espanhol, por aprendizes brasileiros.

O contraponto ao apagamento do clítico seria a sua supergeneralização no *corpus* de aprendizes, cujo índice ficou em 2,1% do total e, conforme González (1994), seria produto da aprendizagem formal (por um efeito mimético) e não da aquisição espontânea.

Por fim, a presença no DEEC de construções impessoais com *SE* não argumental (não observada no *corpus* do espanhol) pode ser atribuída à influência da língua materna, nas quais são atestadas com relativa frequência. O clítico *SE* nessas construções, tal qual parece ocorrer no PB, enfatizaria o caráter humano do sujeito/agente selecionado pelo verbo, embora esse participante não esteja codificado na sentença.

Referências bibliográficas

ARAÚJO JÚNIOR, B. J. **As passivas na produção escrita de brasileiros aprendizes de Espanhol como língua estrangeira**. 2006. 111f. Dissertação (Mestrado em Língua Espanhola). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006.

_____. **Limites precisos ou fronteiras que desaparecem?** As construções impessoais e passivas com o clítico SE/SE no português brasileiro e no espanhol. 2013. 203 f. Tese (Doutorado em Língua Espanhola). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2013.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, Manole, 2004. 410 p.

GONZÁLEZ, N. T. M. **Cadê o pronome? — O gato comeu**. Os pronomes pessoais na aquisição/aprendizagem do espanhol por brasileiros adultos. 1994. 451f. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1994.

HOPPER, P. J.; THOMPSON, S. Transitivity in grammar and discourse. **Language**, v. 56, n.2, p. 251-99, 1980. **crossref** <http://dx.doi.org/10.1353/lan.1980.0017>

KEMMER, S. Middle voice, Transitivity, and the Elaboration of Events. In: FOX, B.; HOPPER, P. J. (eds.) **Voice: Form and Function**. John Benjamins: Amsterdam/Philadelphia, 1994, p. 179-229. **crossref** <http://dx.doi.org/10.1075/tsl.27.09kem>

LANGACKER, R. W. **Foundations of Cognitive Grammar: Descriptive Application**. Stanford: Stanford University Press, 1991, v.2. 590 p.

LYONS, J. **Linguistique générale**. Paris: Larousse, 1970. 384 p.

MALDONADO, R. **A media voz: Problemas conceptuales del clítico se**. México: Universidad Nacional Autónoma de México, 2006. 482 p.

MENDES, R. SP2010 - Construção de uma amostra da fala paulistana. Projeto de Pesquisa apresentado à FAPESP (Proc. no. 2011/09278-6), 2011.

MORENO FERNÁNDEZ, F. et al. **La lengua hablada en Alcalá de Henares. Corpus PRESEEA-ALCALÁ. I. Hablantes de Instrucción Superior.** Alcalá de Henares: Universidad de Alcalá, 2002. CD-ROM.

_____. **La lengua hablada en Alcalá de Henares. Corpus PRESEEA-ALCALÁ. II. Hablantes de Instrucción Media.** Alcalá de Henares: Universidad de Alcalá, 2004. CD-ROM.

Artigo recebido em: 28.09.2014

Artigo aprovado em: 15.12.2014