

COMPOSTO OU SINTAGMA? OBSERVANDO COMPOSTOS NOMINAIS ENDOCÊNTRICOS EM INGLÊS E PORTUGUÊS BRASILEIRO

Julio William Curvelo BARBOSA¹
(Universidade de São Paulo)

Resumo: Seguindo o quadro da Morfologia Distribuída (HALLE; MARANTZ, 1993; MARANTZ, 1997), este trabalho discute o parâmetro proposto em Barbosa (2012), apresentando argumentos semânticos e sintáticos para defender que as expressões N + de + N do português brasileiro sejam estruturalmente idênticas aos compostos N+N do inglês, variando apenas na realização da preposição em PF, o que explicaria as ordens dos núcleos distintas nessas línguas.

Palavras-chave: Composição, Morfologia Distribuída, Parâmetros.

INTRODUÇÃO

O objetivo principal deste artigo é contribuir para o debate da questão do fenômeno da composição como fenômeno “de fronteira”, já que, de acordo com os estudos sobre o assunto, composição sempre foi tomado como o processo de formação de palavras que apresenta características tanto de elementos complexos formados no léxico quanto de elementos formados na sintaxe. É pensando na relação entre léxico e sintaxe que surge a motivação deste trabalho, pois, partindo das definições do modelo teórico da Morfologia Distribuída (HALLE; MARANTZ, 1993, MARANTZ, 1997), elimina-se a distinção entre itens complexos formados no léxico e itens complexos formados na sintaxe. Com esse paradigma teórico em mente, surge a seguinte questão: como lidar com o fenômeno da composição?

A discussão será feita a partir da observação de dados de compostos nominais endocêntricos (compostos-raiz) N+N do inglês (1a) e de expressões N+de+N em português brasileiro (doravante PB) (1b).

- (1) a. *orange juice, coffee cup, leather wallet*
b. *suco de laranja, copo de café, carteira de couro*

A partir da comparação semântica dessas expressões, será apresentada uma hipótese que motive a possibilidade de expressões análogas a *orange juice* e *coffee cup* no PB, e que, ao mesmo tempo, explicita a diferença entre as expressões do inglês e seus equivalentes em PB por meio de um parâmetro semelhante à proposta de Barbosa (2012).

O artigo se segue da seguinte maneira: na seção 2, são apresentadas as características dos nomes compostos definidas nas clássicas sobre compostos na literatura, seja para línguas germânicas, como o inglês, seja para românicas, como o PB. Na seção 3, são apresentados argumentos em favor da hipótese de que N+de+N no PB seja equivalente ao N+N do inglês, com base nas características prototípicas de compostos. A seção 4 apresenta uma análise uniforme para os compostos/sintagmas do PB e do inglês. A conclusão do artigo é vista na seção 5.

¹ Pós-Doutorando- Departamento de Linguística - Universidade de São Paulo. Contato: jbarbosa@usp.br

Julio William Curvelo Barbosa

1. O QUE SÃO COMPOSTOS?

A noção de composto (ou de composição) presente na literatura sobre o assunto, apesar de controversa, delimita algumas propriedades consensuais entre os mais variados pontos de vista teóricos. Neste artigo, serão destacadas: (i) a produtividade, (ii) a interpretação composicional e (iii) a *possibilidade* de interpretação idiossincrática. Essas três propriedades se mostram fundamentais para a discussão empírica das expressões nominais complexas do PB.

1.1. Definições de composição e compostos

Plag (2003) destaca dois tipos de palavras no inglês: as palavras morfológicamente simples – que não podem ser decompostas – como *chair* (*cadeira*) ou *neighbor* (*vizinho*), e as palavras morfológicamente complexas, como *inventor*, *employee* (*empregado*), *apartment building* (*prédio de apartamento*) e *greenhouse* (*estufa*; *lit. verde+casa*), que seriam formadas por derivação (*inventor* = *invent+or*), ou por composição (*greenhouse*, *apartment building*). Uma característica importante da composição para Plag que deve ser levada em conta é que composição é um fenômeno que combina duas raízes.

De maneira semelhante, Basílio (1987) afirma que os processos mais comuns para formação de palavras no PB são a derivação e a composição. Palavras derivadas são aquelas unidas a partir de uma *base* e um *afixo*, tal como *retratista* (*retrato* + *-ista*), *livreiro* (*livro*+*-eiro*), *lavável* (*lava*+*-vel*), *contemplação* (*contempla*+*-ação*), *reler* (*re*+*ler*), e *predispor* (*pré*+*dispor*) (BASÍLIO, 1987: 27). A composição, por sua vez, teria como produto palavras que são a junção de duas bases, como *guarda-chuva* (*guarda*+*chuva*), *luso-brasileiro* (*luso*+*brasileiro*), *sociolinguístico* (*sócio*+*linguístico*) e *agricultura* (*agri*+*cultura*) (BASÍLIO, 1987: 28). Nomes compostos poderiam ser formados tanto com bases presas como *agri*-, quanto por bases livres. Para a autora, a diferença entre a derivação e a composição se dá pelo fato de que cada uma delas satisfaz uma necessidade particular de comunicação. Enquanto a derivação dá conta da expressão de categorias nocionais – nomes de interpretação/significação “simples” –, a composição expressa *combinações particulares*, ou seja, significados idiossincráticos.

Um contraste interessante é feito pela autora no que diz respeito à diferença entre derivação e composição: a derivação gera “expressões comuns e gerais” (BASÍLIO, 1987: 31), enquanto a composição “vai permitir categorizações cada vez mais particulares” (*idem*). Ou seja, para a autora, compostos tendem a se distanciar do significado original de suas partes, pois o composto é a “junção de dois elementos semânticos, de existência independente no léxico, em apenas um elemento lexical” (BASÍLIO, 1987: 31).

Justamente por conta dessa diferença em interpretação, surgem duas questões importantes: compostos são expressões nominais de interpretação livre – isto é, de interpretação composicional – ou são expressões que denotam conteúdos semânticos específicos/idiossincráticos?

Bauke (2011) aponta que os compostos-raiz do alemão podem ser divididos em dois grupos. O primeiro grupo é formado por compostos que não apresentam marcas flexionais entre as raízes, possuem interpretação idiossincrática, sem possibilidade de recursividade (1a). O segundo grupo, por sua vez, é formado por compostos que apresentam marcas flexionais entre as raízes, possuem interpretação composicional, e podem ser recursivos (2b):

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

- (2) a. Landkarte
campo + mapa
‘mapa’
- b. Landsmann
campo + GENITIVO + homem
‘compatriota’ ou ‘homem que ama o campo’ ou ‘homem que defende a preservação do campo’, etc.’
- c. Landeskirche
campo + GEN + igreja
‘igreja nacional’ ou ‘igreja que é associada com o país’ ou ‘igreja que mostra a arquitetura típica do país’, etc.
- d. Länderspiel
country + PL + match
‘partida entre dois times nacionais’ ou ‘jogo que envolve conhecimento entre certos países’ ou ‘jogo que é tipicamente jogado em certos países’, etc.
(BAUKE, 2011: 3)

De acordo com a autora, o fato de compostos como (2a) apresentarem tais propriedades idiossincráticas, além da ausência de marcas morfológicas intervindo entre suas raízes, seria evidência para que tais compostos sejam considerados um produto de formações lexicais, enquanto as expressões de (2b-d), pelo fato de serem expressões composicionais e morfológicamente mais permissivos, seriam formadas na sintaxe.

Olsen (2008) revisa uma série de trabalhos clássicos sobre compostos, e destaca que as seguintes propriedades são recorrentes nos trabalhos que analisam o fenômeno:

- (3) Propriedades típicas da composição, de acordo com Olsen (2008):
- (i) É um processo produtivo;
 - (ii) Pode ser recursivo;
 - (iii) Compostos são vagos e ambíguos;
 - (iv) N+N é seu padrão mais regular;
 - (v) Compostos estão sujeitos a especializações de significado – mas estes significados “devem estar gravados no léxico”.

Seguindo esse ponto de vista, os compostos formados no PB, apontados por Basílio (1987) seriam sempre formados no léxico, assim como os compostos idiossincráticos de línguas como o alemão e o inglês. Essa distinção é importante, já que sempre foi notado na literatura que existe uma diferença entre a produtividade de compostos de línguas germânicas em contraste às línguas românicas. Por outro lado, se uma teoria como a da Morfologia Distribuída (HALLE; MARANTZ, 1993, MARANTZ, 1997, doravante MD) é levada em consideração, a divisão entre léxico e sintaxe como dois lugares distintos onde são geradas expressões complexas se perde, e se torna necessário explicitar onde se dá essa distinção claramente definida a partir da observação empírica dos compostos.

Para a MD, o único componente gerador de expressões complexas, sejam elas palavras ou sentenças, é a sintaxe. Por conta desse fato, a distinção proposta por Bauke (2011) deve ser acomodada dentro desse modelo teórico. Acredita-se que essa acomodação seja possível, e que os dados do PB apresentados ao longo deste artigo permitem mostrar que não apenas a formação de compostos idiossincráticos possa ser explicada no léxico, como a formação de compostos composicionais seja possível em línguas românicas, desde que satisfeitos determinados requerimentos morfofonológicos.

Na próxima seção, serão discutidos os dados do PB presentes em Barbosa (2012). A partir da observação desses dados, e das questões de produtividade, a classificação para

Julio William Curvelo Barbosa

compostos-raiz endocêntricos de Bauke (2011) será questionada, o que permitirá uma estrutura convergente para os compostos do inglês e do PB.

2. O PB E O INGLÊS: ONDE OS COMPOSTOS N+N E SINTAGMAS N+DE+N DO PB SE ASSEMBELHAM

Barbosa (2012), partindo da observação dos dados do PB, discute a cobertura empírica do Parâmetro de Composição de Snyder (1995, 2001). De acordo com o parâmetro proposto por Snyder, línguas que não apresentam compostos N+N endocêntricos de maneira produtiva não apresentariam uma série de predicados complexos, como construções resultativas, verbo+partícula, locativas, perceptuais, dativas e de objeto duplo. Barbosa (2012) questiona a abrangência da proposta de Snyder, e, na busca de estabelecer critérios sintáticos para os tipos de predicados complexos relevantes para o Parâmetro de Composição, sugere que exista uma relação entre os predicados complexos que não ocorrem no inglês e os compostos N+N: a presença da preposição. Construções dativas e de objeto duplo estariam em um par análogo aos compostos N+N e as expressões N+*de*+N do PB, e as suas variantes sem preposição seriam restritas em PB por conta da mesma restrição paramétrica.

Em um estudo longitudinal de aquisição de duas crianças, R e L, entre 1;9 e 4;11, Barbosa & Simioni (2009) apontam que alguns tipos de predicados complexos e expressões nominais complexas surgem na gramática de falantes do PB a partir de 2;0 anos, dentre os quais são destacados os dativos preposicionados:

(4) *Dativos*

- | | |
|--|---------------|
| a. Dá tomada po pintinho. | (R., 2;0.20) |
| b. O Ped(r)inho deu pessinho p(ar)á mim. | (R., 2;2.2) |
| c. Me dá a o(u)tra ponte. | (R., 2;10.14) |
| d. Daí eu ent(r)ego pos meu(s) amiguinho(s). | (L., 4;5.7) |

(5) *Expressões nominais complexas com preposição*

- | | |
|----------------------------------|--------------|
| a. galinha do b[v]id[z]i(...)nho | (R., 2;0.20) |
| b. um pedacinho de bolo | (R., 2;0.20) |
| c. tu(do) caz[r]a de b(r)uxa. | (L., 2;2.28) |
| d. do(r) de ba(rr)iga, mamãe. | (L., 2;3.11) |

(BARBOSA; SIMIONI, 2009)

Com os argumentos de aquisição de Barbosa & Simioni (2009), Barbosa (2012) questiona a impossibilidade de ocorrência de compostos “verdadeiros” no PB: em princípio, seria arbitrário admitir que não exista a possibilidade de combinação entre dois nomes que gerasse uma interpretação composicional simples.

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

2.1. Argumentos empíricos para a equivalência entre N+N germânico e N+de+N românico

Jakobson (1971: 264) afirma que: “*as línguas diferem essencialmente no que elas devem expressar, e não no que elas podem expressar*”. Caso se tome essa afirmação por verdadeira, não existe língua incapaz de gerar nomes compostos – ou, ao menos, expressões equivalentes derivadas por mecanismos computacionais diferentes. Seguindo essa ideia, Barbosa (*op. cit.*) sugere que existam componentes explícitos dentro de uma expressão nominal em PB que licenciem o mesmo tipo de relação que nomes compostos N+N do inglês. Essas expressões, em PB, seriam licenciadas pela preposição *de*, um elemento de natureza relacionadora (HALE; KEYSER, 2002). A preposição *de* seria, portanto, a responsável pelo bloqueio de leituras composicionais em PB, enquanto seu apagamento fonológico, no inglês, permite (ou, talvez, obriga) a formação de compostos N+N a partir da mesma estrutura.

Para confirmar empiricamente essa hipótese, são observados dados que mostram que tanto os dados N+N do inglês quanto os N+*de*+N do PB apresentam as mesmas propriedades gerais de compostos: (i) recursividade, (ii) produtividade (N+N sendo o padrão mais regular), (iii) interpretação variável (vaga/ambígua/composicional), (iv) possibilidade de significado especial (lexicalizado).

Observando os compostos N+*de*+N, é possível notar enorme semelhança nos critérios acima observados. No que diz respeito à (i) recursividade, assim como acontece com os compostos N+N do inglês, o PB é capaz de formar compostos N+*de*+N de enorme recursividade:

(6) [[Bolo de [merengue de limão]] de [[dona de casa] de [cidade de interior]]] é sempre gostoso.

(Cf. ‘(A) countryside city’s housewife lime meringue cake is always tasty’, em inglês)

No caso de (ii) – produtividade/N+N ser o padrão mais regular do inglês –, pode ser citado o fato de que *de* é a palavra mais escrita no PB, especialmente quando precedida por nomes, conforme mostrado em Villavicencio, Finatto & Possamai (2005), baseados no Banco de Português – <http://www2.lael.pucsp.br/corpora/bp/index.htm>, composto então por 223 milhões de palavras.

Em outro estudo, apoiado tanto em PB quanto PE, Davies & Preto-Bay (2008) mostram que *de* é a segunda palavra mais usada no português – aparecendo 1.691.442 vezes em um *corpus* de 20 milhões de palavras, baseado majoritariamente na seção dos anos 1900 do *Corpus do Português* (www.corpusdoportugues.org).

Teixeira (2009), em um estudo sobre a semântica dos nomes compostos, parte de um *corpus* de dez edições da revista *National Geographic*. Ao analisar os dados a fim de observar os correspondentes de tradução dos compostos em seu *corpus*, a autora observa que

[...] a construção N de N é de fato a mais utilizada por tradutores humanos para expressar a relação entre os elementos de um composto em língua portuguesa, totalizando 91 (incluindo as expressões com artigos) das 165 ocorrências analisadas, conforme dados da tabela [...] (TEIXEIRA, 2009: 135).

Além da produtividade, Teixeira ressalta o fato de que “*devido ao seu aspecto polissêmico, a preposição de pode expressar qualquer uma das relações analisadas nesta seção: função, origem, material, parte,*

conteúdo, local, tempo e posse” (TEIXEIRA, 2009: 135). Como ilustração da afirmação feita por Teixeira, pode ser citado um dicionário de compostos do inglês de conteúdo aberto, o *Wiktionary*². Nele, existem 7.148 entradas (último acesso em 18/05/2012), enquanto na versão em português, o *Wikcionário*³ existem apenas 217 entradas⁴. Destas 217 entradas, 74 (34,1%) são formadas com a preposição *de*. Ou seja, é inegável a produtividade de expressões com a preposição *de*, seja na forma de compostos lexicalizados, seja na sua forma “sintática”.

Gomes (2009) também fala da abrangência da preposição *de* nos compostos. A autora vai contra a proposta de Neves (2000, *apud* GOMES, 2009), que afirma que a preposição *de* estabelece uma relação no tempo e no espaço; Gomes aponta que em compostos como *jantar do sábado* ou *mesa do fundo*, a preposição “de” está ali apenas como elemento de ligação entre o antecedente, e as referências de tempo ou espaço são dadas por *sábado* e *fundo*, respectivamente. A autora conclui que “... *fora do contexto, o significado da preposição “de” é imprevisível, pois esta é extremamente dependente dos elementos que a antecedem e sucedem.*” (GOMES, 2009: 109).

O fato de a preposição *de* apresentar um caráter genérico é um argumento importante em favor do item (iii) (interpretação variável (vaga/ambígua/composicional)). Apesar de os itens (iii) e (iv) dizerem respeito a propriedades semânticas opostas – já que é impossível ser vago e apresentar um significado especializado ao mesmo tempo –, a relação entre composicionalidade e lexicalização é muito próxima quando se observa a formação de compostos pelo panorama da Morfologia Distribuída, e o papel da preposição *de* se revelará crucial para explicar as diferenças entre PB e inglês.

Pensando na questão da lexicalização das expressões com *de* em contraposição à liberdade de formação de compostos de interpretação livre/composicional no inglês, é possível ver que a preposição *de* mostra o mesmo tipo de ambiguidade que um composto quando em contraste com uma expressão com a preposição *para* no PB (7). Ou seja, expressões em PB com *de* também possuem informações implícitas, dependentes de contexto.

- [illegible]

² <http://en.wiktionary.org/wiki/Category:English_compound_words>.

³ <[http://pt.wiktionary.org/wiki/Categoria:Substantivo_composto_\(Português\)](http://pt.wiktionary.org/wiki/Categoria:Substantivo_composto_(Português))>.

⁴ É óbvio que essas listas não correspondem à realidade lexical das línguas, nem refletem a frequência da formação de compostos em inglês ou em português. Porém, dicionários também apresentam uma infinidade de termos que fogem do conhecimento lexical da grande maioria dos falantes de qualquer língua, e, mesmo o mais culto dos falantes não é capaz de saber todas as acepções de todos os nomes compostos de sua língua. Assim sendo, a utilização dos dados aqui mencionados são por fator de ilustração, já que qualquer tentativa de formação de um *corpus* de compostos de inglês e PB seria tarefa que foge do escopo do trabalho aqui delineado.

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

O mesmo pode ser dito dos dados de (8) e (9)(9). Dependendo do elemento que modifica o núcleo das expressões com *de* no PB, surge uma interpretação idiomática, que depende de conhecimento extralinguístico para sua interpretação, assim como a palavra *pancake* (panqueca), cuja interpretação não corresponde à interpretação de suas partes, pode ser decomposta em *pan* (frigideira) + *cake* (bolo) = (lit. ‘bolo de frigideira’), mas *carrot cake* (bolo de cenoura) mantém sua interpretação literal. Um argumento interessante pode ser notado em (8). Enquanto o PB faz uma distinção entre (8b) e (8c), o mesmo nome composto denota os dois objetos no inglês (8e-f). No caso dos contrastes entre (9b-c) e (9e-f), a expressão do PB em (9c) é equivalente a um único item lexical do inglês.

(8) a. *colher de brinquedo*

b. *colher de madeira*

c. *colher de pau*

d. *toy spoon*
brinquedo colher
“colher de brinquedo”

e. *wooden spoon*
“madeiral” colher
“colher de madeira”

f. *wooden spoon*
“madeiral” colher
“colher de pau”

(BARBOSA, 2012: 116)

(9) a. *cara de boneca*

b. *cara de madeira*

c. *cara de pau*

d. *doll face*
boneca rosto
“cara de boneca”

e. *wooden face*
“madeiral” rosto
“cara de madeira”

f. *chutzpah*⁵
“cara de pau”

(BARBOSA, 2012: 115)

Além das expressões N+N, outras combinações de “raiz+raiz” possuem equivalentes em PB por intermédio da preposição *de*; compostos adjetivais como (10), por exemplo, aparentam ter

⁵ A palavra *chutzpah* advém de um empréstimo do hebraico, cujo significado na língua de origem é *audácia ou insolência*.

Julio William Curvelo Barbosa

as mesmas propriedades quando convertidas em expressões $A + de + A$. Os adjetivos que são núcleos da expressão (à direita, no inglês; à esquerda, no PB) recebem um tipo de modificação de grau, equivalente a um advérbio/locução adverbial de modo (cf. (10(10)a"-b") e (10(10)a'''-b''')). O interessante é que este tipo de composição é extremamente distinto da contraparte sintática em (11) (abaixo), que também usa a preposição *de* (*of*, no inglês); por outro lado, a expressão em colchetes indica o complemento do adjetivo, e não o grau deste.

- (10) a. I'm dead tired.
 a'. Eu estou morto de cansado/cansaço.
 a''. I'm extremely tired.
 a'''. Eu estou extremamente cansado.
- b. The kids are crazy hungry.
 b'. As crianças estão doidas de fome.
 b''. The kids are very hungry.
 b'''. As crianças estão *com muita* fome/muito famintas.
- c. The students are zombie sleepy.
 c'. Os alunos tão zumbis de sono/mortos de sono/acabados de sono.
 c''. The students are quite sleepy.
 c'''. Os alunos estão com muito sono/muito sonolentos. (BARBOSA, 2012: 115)
- (11) a. I am tired [of waiting for the bus].
 a'. Eu estou cansado [de esperar pelo ônibus].
- b. Her parents are proud [of her good grades].
 b'. Os pais dela estão orgulhosos [de suas notas boas]. (BARBOSA, 2012: 115)

Pelos dados em (12), é possível notar que a interpretação livre dos compostos/sintagmas $N+de+N$ pode ser representada por dois traços mais básicos: [locativo] e [posse]. Barbosa (2012) mostra que esses traços também estão presentes nas leituras prototípicas de construções dativas e de objeto duplo, e propõe que a distribuição da realização fonológica para cada um desses traços se dê conforme as tabelas em (13) e (14), abaixo:

- (12) a. carrot cake = a cake made of/with carrots
 'bolo de cenoura = um bolo feito de/com cenouras'
- b. wooden house = a house made of/with wood
 'casa de madeira = uma casa feita de/com madeira'
- c. amusement park = a park with amusing rides...
 'parque de diversões = um parque com brinquedos divertidos'
- d. London girl = garota (que está sempre/vem) em/de Londres
- e. USP student = aluno (que estuda/vem) na/da USP
- f. freshwater fish = peixe (que vive/é encontrado) em água doce
 (BARBOSA, 2012: 165-166)

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

(13) Possibilidades de combinação dos traços [posse] e [locativo] em inglês

INGLÊS			
v	v	n	n
[causa]	[causa]	[entidade]	[entidade]
+	+	+	+
F	F	F	F
[posse] (∅)	[locativo] (to)	[locativo] (∅) [posse] ('s)	[locativo] (from) [posse] (of)
=	=	=	=
Objeto duplo	Dativo	"Compound"	Compostos preposicionados

(BARBOSA, 2012: 167)

(14) Possibilidades de combinação dos traços [posse] e [locativo] em PB

PB/PBM			
v	v	n	n
[causa]	[causa]	[entidade]	[entidade]
+	+	+	+
F	F	F	F
	[locativo] (a/para) [posse] (a/para)		[locativo] (de) [posse] (de)
=	=	=	=
Impossível	Dativo	Impossível	Compostos preposicionados

(BARBOSA, 2012: 167)

De acordo com os dados apresentados até aqui, surge um problema quanto à classificação do que é composto, seja em PB, seja em inglês; foi visto acima que os compostos do PB formados por N+*de*+N são de caráter produtivo, o que evidencia sua capacidade de interpretação não-lexicalizada, assim como os compostos N+N, representantes característicos do processo de composição do inglês. Logo, é possível afirmar que N+*de*+N seria a *contraparte românica* de N+N, e que esse tipo de expressão, assim como as expressões do inglês que possuem interpretação composicional, seriam *sintagmas*, e não compostos. A diferença é que, enquanto compostos N+N do inglês são representantes tanto de formas composicionais livres quanto de formações com significação especial, o PB só forma N+N em casos de leitura idiossincrática. Apesar disso, é possível que expressões N+*de*+N também apresentem compostos/sintagmas de interpretação idiossincrática.

Por outro lado, as expressões N+*de*+N do PB representam uma grande porcentagem do número de compostos de significado não-composicional. Da mesma maneira, os compostos do inglês de forma N+N são altamente produtivos, e são grande parte das expressões nominais cristalizadas.

Um contraponto que poderia ser argumentado é o seguinte: Snyder (1995) afirma que um composto do francês – uma língua românica, marcada negativamente para o Parâmetro de

Julio William Curvelo Barbosa

Composição assim como PB – como *tasse à café* (copo de café) é o equivalente sintático de *coffee cup*, e, ao contrário do inglês, a expressão do francês não pode ser considerada uma unidade no nível da palavra. Logo, expressões N+*de*+N seriam produto de formação sintática, exceto quando sua união resulta em uma palavra com novo significado. Porém, os dados acima não permitem ignorar a semelhança semântica entre compostos N+*de*+N do PB e os compostos N+N do inglês, o que faz crer que exista uma semelhança sintática entre essas expressões, em algum momento da sua formação.

Por consequência da distinção entre “interpretação especial dos compostos surge no léxico” *versus* “interpretação composicional é feita na sintaxe”, a distinção entre N+N, formado no léxico e N+*de*+N, formado na sintaxe, poderia explicar a formação de compostos do inglês como paralela à formação de compostos N + *de* + N do PB. Se os compostos do inglês são interpretados como uma palavra morfológicamente complexa, devem ser formados no léxico. Apesar disso, a produtividade de expressões N+N é enorme, fato geralmente correspondente às expressões sintáticas, e não às lexicais.

Assim sendo, ao invés de afirmar que compostos N+N interpretados de maneira composicional são formados na sintaxe – o que seria um problema para a distinção léxico/sintaxe –, será proposto que os compostos do inglês sejam gerados a partir da mesma estrutura que expressões N+*de*+N, e outros processos fazem com que o produto final seja N+N; essas questões serão explicadas com base no modelo da Morfologia Distribuída, em especial a noção de Marantz (2001) para a definição de palavra, e para determinar as questões de interpretação especial nos compostos. Na próxima seção, o trabalho de Marantz é exposto, de modo a auxiliar na análise dos compostos do PB e do inglês.

2.2. Idiomatidade estrutural: Marantz (2001)

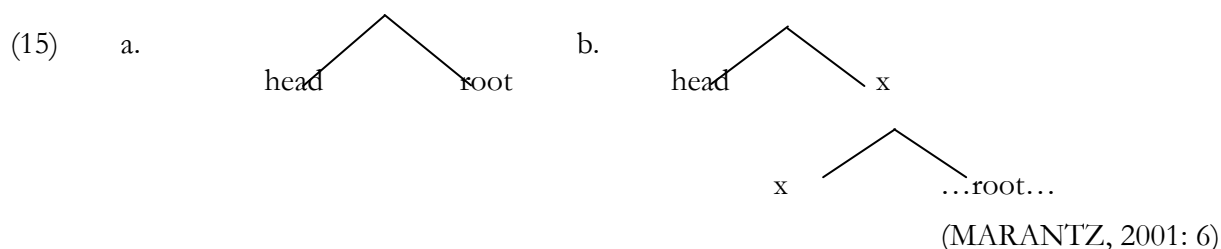
Em um texto que defende a existência de apenas um componente computacional que seja responsável pela formação de sentenças e palavras, Marantz (2001), dentro do panorama da Morfologia Distribuída, propõe que a noção clássica de “dois lugares” para a formação de objetos computacionalmente complexos, atribuídos previamente ao léxico (para a formação de palavras) e à sintaxe (para a formação de sentenças) seja substituída por composição gerada apenas na sintaxe, a partir das relações estruturais – posições na árvore sintática – e derivacionais dos itens presentes na estrutura sintática, envolvendo, segundo o autor, domínios cíclicos. Dessa forma, Marantz faz as seguintes afirmações:

A uniformidade da morfofonologia advém da natureza interpretativa da [própria] morfofonologia, que uniformemente forma a sintaxe.

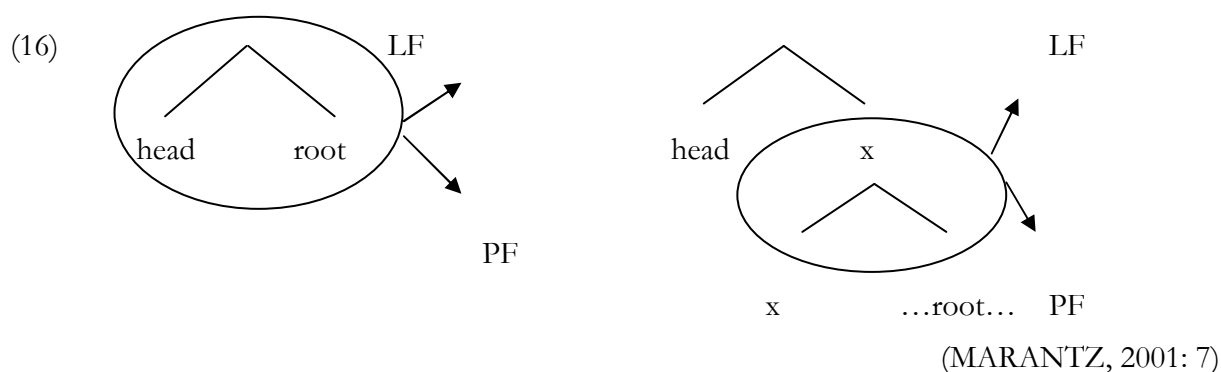
A uniformidade da composicionalidade advém do fato de a sintaxe realizar todas as operações de concatenação [merger], incluindo aquelas entre morfemas nos limites da palavra. (MARANTZ, 2001: 6, tradução minha).

Seguindo essa proposta, Marantz propõe que existam duas maneiras de formar uma palavra derivacionalmente. Na primeira, pode-se unir morfemas (*head*, cf. (15a), abaixo) a uma raiz, que, por sua vez, será então categorizada (nome, verbo, adjetivo, etc.). A outra maneira de formar palavras é modificando essa raiz já categorizada, onde um elemento x (n , a , v) intervém entre a raiz e esse morfema formador de palavra (*head* em (15b)).

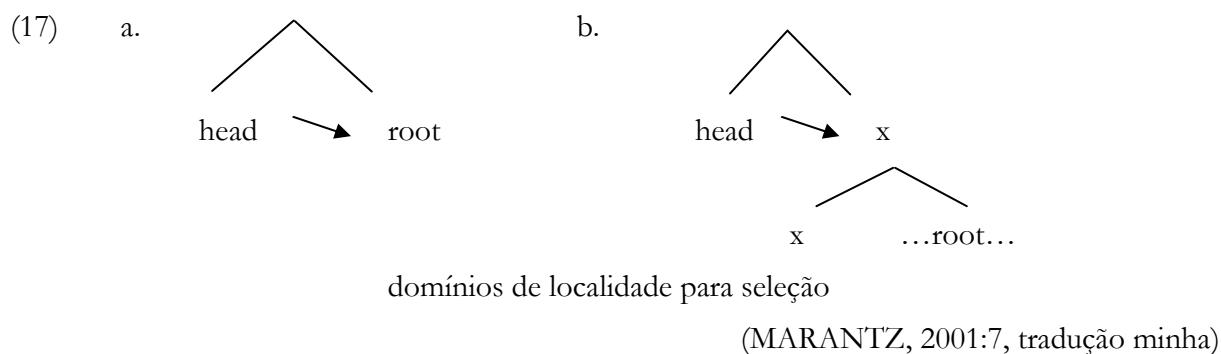
Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro



Marantz (2001) afirma que esses núcleos (x , correspondentes aos núcleos n , v , ou a) determinam a “borda” de um domínio cíclico, ou uma fase, nos termos de Chomsky (1999). Ou seja, morfemas que se unem à raiz antes da inserção desse núcleo x ainda estão no domínio da palavra (15a), enquanto morfemas que não são c-comandados por x estão fora desse domínio (15b). A estrutura formada pela união da raiz com esse núcleo x é enviada para LF e PF para interpretação fonológica e semântica, e o significado da raiz no contexto de x “zinho” é “negociado”, nos termos de Marantz, usando conhecimento Enciclopédico (lista 3; cf. seção 1.4., acima). Assim, núcleos que se unem a x “zinho” selecionariam como complemento uma estrutura em que o significado (e a pronúncia) já teriam sido negociados (16):



Marantz argumenta que, “se um núcleo se concatena fora de x “zinho”, ele vê os traços de x localmente, mas não os traços, propriedades, ou identidade da raiz concatenada com x ”⁶. Nesse caso, as propriedades de seleção de um núcleo são satisfeitas pelos traços de x , e não pelas propriedades da raiz, que são, segundo Marantz, “idiossincráticas à língua e ao falante individual” (17b). Se o núcleo se une à raiz, os requerimentos de seleção passam a ser satisfeitos pelas propriedades idiossincráticas da raiz (17a):



⁶ MARANTZ, 2001: 7.

Julio William Curvelo Barbosa

Com a proposta de Marantz, tem-se um instrumental teórico que explicaria a noção de “dois lugares” para formação de palavras. As propriedades clássicas de palavras “formadas no léxico” e palavras “formadas na sintaxe” se dão pela junção dos morfemas derivacionais no domínio de localidade relevante para a sua interpretação.

2.3. Questões de intervenção: definitude e modificação afetando idiomaticidade

Caso sejam adotadas as restrições estruturais propostas por Marantz, ficam explicadas, sem a necessidade de um léxico, as oposições (i) *interno vs. externo*, (ii) *semiprodutivo vs. produtivo*, (iii) *tem a ver com a semântica das raízes vs. tem a ver com a estrutura argumental sintática*, (iv) *é associado com significados especiais vs. é associado com significados previsíveis*. Dizer que essas relações estão diretamente ligadas aos núcleos categorizadores de raízes explicará a relação entre os compostos do inglês e do PB, apesar de suas formas superficiais distintas. Essa relação será mostrada na próxima seção.

Pensando ainda nos domínios de interpretação, é possível fazer um paralelo entre as expressões N+*de*+N que apresentam artigo definido junto à preposição *de* e as marcas de flexão dos compostos de leitura composicional do alemão em (2b-d), acima. Se a proposta de Marantz estiver correta, parece que existe uma necessidade de domínio fechado para que a interpretação do composto seja idiomática.

- (18) a. homem do(s) sapo(s) = *frog man*
 b. ?homem de sapo = *frog man*
 c. cara de sapo = *frog face*
 d. cara do sapo = *frog guy*
 e. Ele é um/*o filho da mãe = *‘He is a bastard’*
 f. Ele é o filhinho da/*de⁷ mamãe = *‘He is mommy’s boy’*
 g. Ele é um filhinho de/*da mamãe = *‘He is a mommy’s boy’*

De acordo com a distribuição das interpretações vistas acima, a interpretação idiomática das expressões N+*de*+N depende que o complemento da preposição não possua nenhuma marca de definitude. Da mesma maneira, Lee (1997) mostra que o plural de compostos que o autor chama de pós-lexicais, como os em (19), só permite modificação do núcleo – o elemento à esquerda – e que qualquer outra marcação de plural é agramatical (20):

- (19) fim de semana, pé-de-moleque (LEE, 1997, (4c))
 (20) *fins de semanas, fim de semanas (LEE, 1997, (17d))

Lee (comunicação pessoal) também aponta que apenas modificações derivacionais podem intervir nos compostos, como diminutivos (*pezinho de moleque*, mas **pé-de-molequinho*), mas que intervenções sintáticas, como de adjetivos (**pé-grande-de-moleque*) são impossíveis. Curiosamente, uma expressão modificada como *pé de moleque grande* é ambígua entre uma leitura de *pé-de-moleque* como composto idiossincrático e *pé de moleque*, uma expressão de leitura

⁷ Em dialetos do PB em que o artigo definido não é obrigatório em frente a um nome próprio, a ausência de concordância é gramatical.

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

composicional⁸. Ou seja, é inegável que existam dois tipos de formação de compostos. Porém, essa diferença pode ser atribuída ao tipo de formação sintática de cada um desses compostos. Essas diferenças serão mostradas a seguir.

3. ESTRUTURAS COMUNS, RESULTADOS DIFERENTES: IDIOMATICIDADE VERSUS COMPOSICIONALIDADE EM PB E INGLÊS

A formação de compostos sempre foi um objeto de debate dentro das teorias sintáticas e morfológicas acerca da estruturação do léxico e dos fenômenos que nele podem (ou não) ocorrer. Porém, se compostos e expressões N+*de*+N – consideradas como expressões derivadas sintaticamente – são, de fato, expressões sintáticas que expressam o mesmo conteúdo semântico, qual seria a necessidade de pensar que nomes compostos sejam de fato, formados em um lugar diferente de sentenças?

Já não é novidade na teoria gerativista a ideia de que exista um paralelismo entre nomes e sentenças (*e.g.*, ABNEY, 1987, *inter alia*). Conforme mencionado na seção 2, a teoria da Morfologia Distribuída propõe que qualquer formação que necessite lidar com operações computacionais para formar estruturas – sejam elas estruturas de palavras ou de sentenças – deva ser gerada em um único componente gerativo – neste caso, o componente sintático. A partir dessa propriedade do modelo teórico adotado, a consequência é que, apesar de possuírem propriedades ditas “do léxico”, todos os tipos de compostos sejam gerados na sintaxe⁹. Um parâmetro explicaria a diferença na estruturação sintática dos compostos e a variação superficial vista entre PB e inglês.

É possível observar que, nos compostos do inglês, apenas aqueles que são classificados como endocêntricos respeitam a *right-hand rule* (WILLIAMS, 1981), que afirma que o núcleo dos compostos deve sempre estar à direita. No PB, estes compostos nunca aparecem lexicalizados com a preposição *de*, o que faz ser possível afirmar que os compostos de formação endocêntrica nunca possuem interpretação composicional, e são sempre produto de lexicalização de significado. Assim sendo, apenas nos casos endocêntricos (em que um dos membros do composto é o núcleo), o processo de composição opera livremente no inglês, o que explicaria a ordem distinta entre núcleos ((21a) e (21c) e não-núcleos ((21b),(21d)):

(21) Comparação entre compostos do inglês e PB, divididos por relação sintática:

a. Relações núcleo-modificador (endocêntricas)

- (i) student film society = sociedade de filme de estudante(s)

b. Relações argumento-predicado (exocêntricas)

- | | |
|------------------|-------------------|
| (ii) pick pocket | (v) fura-olho |
| (iii) cut-throat | (vi) quebra-galho |

⁸ É importante notar que existe uma interação entre a definitude do DP como um todo e a possibilidade de marcação de definitude do elemento que vem depois da preposição:

- i. Ele é um/*o filho da mãe.
- ii. Ele é *um/o filho da mãe que bateu no meu carro.
- iii. Ele é o filhinho da/*de mamãe.
- iv. Ele é um filhinho de/*da mamãe.

⁹ Para uma análise do fenômeno de composição baseada na Morfologia Distribuída, cf. Minussi (2008).

Julio William Curvelo Barbosa

(iv) lazy-bones

(vii) risca-faca

c. Relações argumento-predicado (endocêntricas)

(viii) truck driver er = motorista de caminhão

(ix) jet pilot o = piloto de jato

d. Relações apositivas (exocêntricas)

(x) mother-child (relationship)

(xiii) (relação) mãe (e) filho

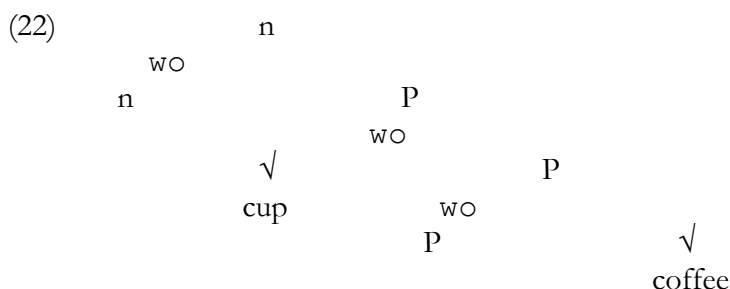
(xi) sofa-bed

(xiv) sofá-cama

(xii) bed and breakfast

(xv) entra (e) sai

A partir da noção de DiSciullo (2005), que propõe que toda relação de compostos apresente uma projeção funcional mediando as duas raízes, será postulado que a projeção P, a mediadora dos elementos do composto, seleciona uma raiz como especificador e uma raiz como complemento. No caso de compostos de interpretação idiossincrática/“idiomática”, a raiz na posição de complemento é selecionada diretamente por P, como em (22):



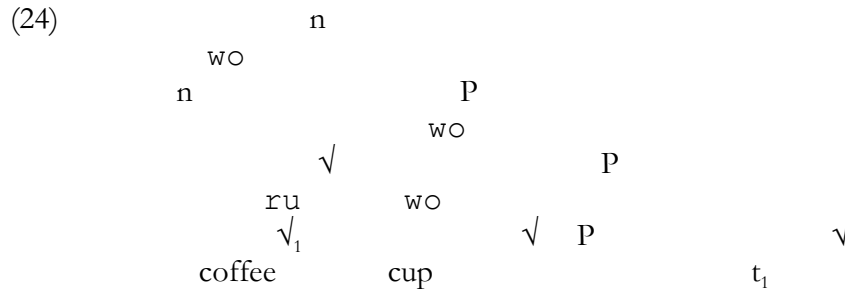
Essa estrutura é a maneira como um composto como *coffee cup* ou um sintagma *cup of coffee* é enviado para LF, após *spell-out*. Para derivar uma forma de superfície N+N, é postulada a regra de deslocamento compulsório (23), que gera a estrutura (24), compatível com a ordem linear do inglês¹⁰.

(23) *Regra de Deslocamento Compulsório*

Um elemento ocupando a posição [Comp, P] deve se concatenar a [Spec, P], sempre que F não tiver conteúdo fonológico.

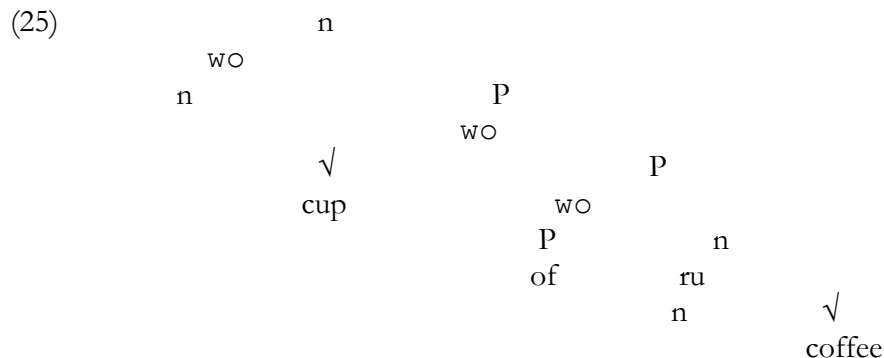
¹⁰ Para explicar a linearização correta apesar da ausência de c-comando assimétrico entre as raízes, é seguida a proposta de Embick (2007). Para o autor, linearização é baseada em propriedades de *Deslocamento Local*, no qual elementos que pertencem ao grupo chamado de *subpalavra* (elementos morfológicamente dependentes) sofrem um processo de linearização inverso; dentro dessas subpalavras, o elemento mais baixo da relação de c-comando é aquele que precederá o elemento que o c-comanda.

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro



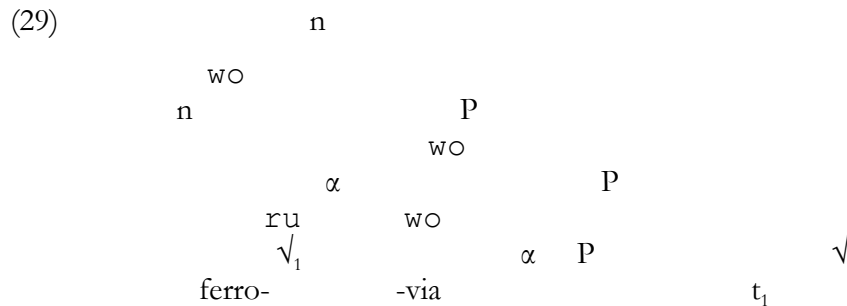
Na estrutura em (22) (acima), os elementos se relacionam graças ao conteúdo relacionador da preposição. De acordo com Hale & Keyser (2002), preposições são elementos *birrelacionais* por natureza, pois “especificam uma relação (espacial, temporal, ou outra) entre duas entidades (ou dois eventos, circunstâncias, etc.)”. (HALE; KEYSER, 2002: 8). Este tipo de relação estrutural é totalmente compatível com as propriedades semânticas dos compostos que foram mostradas nas seções anteriores, frequentemente tidas como vagas e abrangentes.

No caso do sintagma *cup of coffee*, do inglês, a preposição se faz presente, e o complemento de P é uma raiz já categorizada (cf. (25), abaixo). Esse tipo de expressão apresenta uma leitura composicional, e além de uma interpretação idiossincrática, (uma medida de café), pode ser interpretada como um objeto utilizado para beber café (a mesma interpretação idiossincrática de *coffee cup*). Ou seja, parece que a interpretação idiomática presente pelo “fechamento da fase” proposta por Marantz (2001) é opcional nos casos em que o núcleo mediador da relação entre as raízes é realizado fonologicamente. Isso explicaria as questões de intervenção de modificadores em compostos de interpretação composicional em PB, inglês ou alemão, já que a raiz em Spec, P é livre para fazer sua negociação de significado independentemente da raiz que é selecionada como complemento de P.



Como a preposição é exigida pela raiz categorizada por questões de marcação de Caso, a estrutura em (25) não poderia derivar um composto como **coffee cup of*, já que deslocamento compulsório não se aplicaria:

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro



Dessa forma, se explica a produtividade dos compostos N+*de*+N nas expressões “lexicalizadas”, ao mesmo tempo em que se explica a produtividade e interpretação vaga/composicional permitida por tal formação estrutural. Fica por ser explicada a questão da interação entre definitude e as interpretações idiossincráticas e composicionais, como nos casos em (18), e na nota 6. Essas questões serão retomadas em trabalhos posteriores.

Um argumento adicional para a motivação da preposição mediando as expressões nominais do PB e do inglês pode ser derivado dos requerimentos do nó terminal P do PB, em contraste ao inglês. Em ambas as línguas, a inserção vocabular (*i.e.*, de conteúdo fonológico) correspondente aos traços do nó sintático P parece ser diretamente relacionada à sua capacidade de satisfazer os traços de Caso de seu complemento na derivação sintática. Qualquer raiz na posição de especificador de P não necessitaria de Caso até ser categorizada, daí sua possibilidade de formar verbos como *to babysit* em inglês (‘cuidar de crianças’, lit. *bebê sentar*), expressão correlacionada com *baby-sitter* (‘babá’), já que o “composto” $\sqrt{+}\sqrt{+}$ é visto como um único elemento dependente de Caso.

A variação entre inglês e PB se dá pela impossibilidade de atribuição de Caso para os dois elementos do sintagma complexo em PB, em analogia às nominalizações passivas (FIENGO, 1979, *apud* JAEGLI, 1986):

- (30)
- | | | |
|---|----------------|------------------|
| a. *irregularities' | acknowledgment | by the senators |
| irregularidades+Gen | reconhecimento | por os senadores |
| 'reconhecimento de irregularidades pelos senadores' | | |
| b. the acknowledgment of irregularities by the senators | | |
| c. *laranja suco | | |
| c' suco de laranja | | |
| d. *irregularidades reconhecimento pelos senadores | | |
| d' reconhecimento de irregularidades pelos senadores | | |

Com a análise delineada acima, acredita-se ter uma proposta uniforme para os nomes compostos, capaz de mostrar que a variação de PB e inglês é produto de variação paramétrica.

Julio William Curvelo Barbosa

CONCLUSÃO

Neste trabalho, foi mostrado que expressões nominais complexas tanto em línguas românicas (como o PB) quanto em línguas germânicas (como o inglês) são passíveis de expressões sintáticas semelhantes, derivadas a partir de uma restrição paramétrica. A categorização da raiz complemento de P (como um DP) permite a negociação de significado (lista 3) composicional seja em inglês/alemão quanto em PB. Compostos com leitura idiossincrática são um subproduto da reinterpretação de estruturas de leitura originalmente composicional, e a negociação de significado parece depender de outros fatores além da categorização da raiz (como definitude, por exemplo).

Se a hipótese aqui levantada estiver correta, composição, assim como todos os processos de formação de palavras, é um fenômeno sintático. Essa é uma condição desejável dentro do modelo da Morfologia Distribuída, e a adequação empírica da proposta desenvolvida neste artigo parece permitir tais afirmações.

REFERÊNCIAS

- ABNEY, S. *The Noun Phrase in Its Sentential Aspect*. MIT Ph. D. Dissertation. Cambridge. 1987.
- BARBOSA, J. *Variação paramétrica em predicados complexos e nomes compostos: um estudo translinguístico*. Tese (Doutorado em Linguística) – Universidade de São Paulo, São Paulo, 222f. 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8139/tde-06032013-094902/pt-br.php>> Acesso em: 30-07-2013.
- BARBOSA, J.; SIMIONI, L. Aquisição do Parâmetro de Composicionalidade: uma análise translinguística. In: NAVES, R. R.; SALLES, H. M. M. L. (eds.) *Estudos Formais da Gramática das Línguas Naturais*. Brasília: Cânone Editorial, 2009.
- BASILIO, M. *Teoria Lexical*. São Paulo: Ática, 1987.
- BAUER, L. Compounds and minor word-formation types. In: AARTS, B.; MCMAHON, A. (Eds.). *The Handbook of English Linguistics*, Malden, MA: Blackwell, p. 483-506, 2006.
- BAUKE, L. S. Phase theory, linearization and antisymmetry in morphology. *Language at the University of Essex (LANGUE) 2010 Proceedings*, p. 1-15, 2011.
- DAVIES, M.; PRETO-BAY, A. M. R. *A Frequency Dictionary of Portuguese*. Bookseller: The Book Depository. Guernsey, GY, United Kingdom. Routledge, 2008.
- DI SCIULLO, A. M. Decomposing Compounds. *Skase Journal of theoretical linguistics*. Vol. 2, n. 3, p. 14-33, 2005.
- EMBICK, David. Linearization and Local Dislocation: Derivational mechanics and interactions. In: *Linguistic Analysis*. 33: 3-4. 2007. Disponível em: <<http://www.ling.upenn.edu/~embick/lin.pdf>>. Acesso em: 04-09-2012.
- GOMES, T. V. *Os nomes compostos em português*. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 74 f. 2005. Disponível em: <http://www.maxwell.lambda.ele.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=7461@1>. Acesso em: 06-11-2012.
- HALE, K.; KEYSER, S. J. Prolegomenon to a theory of argument structure. *Linguistic inquiry monographs*. v. 39. Massachusetts: MIT Press, 2002.

Composto ou sintagma? Observando compostos nominais endocêntricos em inglês e português brasileiro

- HALLE, M.; MARANTZ, A. Distributed morphology and the pieces of inflection. In: HALLE, M.; MARANTZ, A.; KEYSER, S. J. (eds.). *The view from building 20: Essays in Linguistics in Honor of Sylvian Bromberger*. Cambridge, MA: MIT Press. p. 111-176, 1993.
- JAEGGLI, O.A. Passive. *Linguistic Inquiry*, v. 17, n. 4, p. 587-622, 1986.
- JAKOBSON, R. *Selected Writings*. vol. II, The Hague: Mouton. p. 260-266. 1971.
- LEE, S. H. Sobre os Compostos do PB. *DELTA*, São Paulo, v. 13, n. 1, Fev. 1997. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44501997000100002&lng=en&nrm=iso>. Acesso em 15-07-2013.
- MARANTZ, A. *Words*. *West Coast Conference on Formal Linguistics*. University of Southern California, Los Angeles. 2001. Disponível em: <<http://web.mit.edu/marantz/Public/EALING/WordsWCCFL.pdf>>. Acesso em: 01-08-2010.
- MARANTZ, A..No escape from syntax: don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers*. v. 4, n. 2, 1997.
- MINUSSI, R. D. *A relação entre Caso e definitude no hebraico: o Construct State e a Marcação Diferencial de Objeto*. Dissertação (Mestrado em Linguística) – Universidade de São Paulo, São Paulo, 2012. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/8/8139/tde-30092008-152504/pt-br.php>>. 2008. Acesso em 30-07-2013.
- OLSEN, S. Semantics of Compounds. In: MAIENBORN, C.; HEUSINGER, K. V.; PORTNER, P. (Eds.). *Semantics: An International Handbook of Natural Language Meaning* = Handbücher zur Sprach- und Kommunikationswissenschaft, Berlin: Mouton de Gruyter. 2008.
- PLAG, I. *Word-formation in English*. *Cambridge Textbooks in Linguistics*. Cambridge: Cambridge University Press. 2003.
- SNYDER, W. B. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77. p. 324-342, 2001.
- SNYDER, W. B. *Language Acquisition and Language Variation: The Role of Morphology*. Ph.D. dissertation, MIT. Cambridge, MA: MIT Press, 1995.
- TEIXEIRA, L. F. *A semântica dos compostos nominais - um estudo de corpus paralelo inglês/português*. 2009. Dissertação (Mestrado em Linguística Aplicada) – Universidade do Vale do Rio dos Sinos, São Leopoldo, 2009. Disponível em: <http://bdtd.unisinos.br/tde_arquivos/7/TDE-2009-07-23T110437Z-799/Publico/TeixeiraLilianFigueiroLinguistica.pdf>. Acesso em: 06-11-2012.
- WILLIAMS, E. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12. p. 245-274. 1981.

Julio William Curvelo Barbosa

COMPOUND OR PHRASE? ANALIZING NOMINAL ENDOCENTRIC COMPOUNDS IN ENGLISH AND BRAZILIAN PORTUGUESE

Abstract: *Following the Distributed Morphology framework (Halle & Marantz, 1993, Marantz, 1997), this work discuss the parameter advanced in Barbosa (2012), presenting syntactic and semantic arguments in order to defend that N+de+N expressions in Brazilian Portuguese are structurally identical to the N+N compounds from English, differing only in the realization of the preposition at PF, what would explain the distinct head orders in these languages.*

Keywords: *Compounding, Distributed Morphology, Parameters.*