

FORMAÇÃO DE COMPOSTOS ENDÓGENOS NO PORTUGUÊS BRASILEIRO: UMA INVESTIGAÇÃO PSICOLINGÜÍSTICA

Rui ROTHE-NEVES¹

- **RESUMO:** Este estudo investigou quais especificadores ocorrem mais frequentemente na formação de compostos endógenos de estrutura N-de-N cujo núcleo é um "objeto estático não-simbólico". Dados de falantes nativos (n = 40) do Português Brasileiro (PB) foram levantados por meio de um teste de preenchimento (*completion test*), construído a partir de uma análise semântica de 62 compostos endógenos, conforme Frederiksen (1975). Itens de alta expectativa serviram de distratores. O foco do estudo foram itens de baixa expectativa (n = 898). Estes apresentaram, em sua maioria (98,3%), especificadores com as funções de "parte", "objetivo" e "objeto não relacionado". Discutiu-se a distinção entre as duas últimas funções, à luz da distinção entre palavras possíveis e reais. Apresentou-se, para futura investigação, a hipótese de que o traço semântico "continente" seja selecionado na formação de endógenos com o mesmo núcleo.
- **PALAVRAS-CHAVE:** Formação de palavras; compostos nominais; psicolingüística.

Introdução

Entende-se por compostos nominais aqueles itens lexicais que apresentam dois itens nominais (N ou A) formando uma unidade semântica e sintática. Este estudo concentrou-se nos compostos do Português Brasileiro (PB) que apresentam estrutura N-de-N, doravante chamados

¹ Departamento de Psicologia – FAFICH/Universidade Federal de Minas Gerais – 31270-901 – Belo Horizonte – MG.

sujeitos normais, para obter dados de produção de compostos endógenos de estrutura N-de-N. Os compostos novos produzidos foram confrontados com uma sistematização prévia, em bases semânticas, do ambiente possível de ocorrência dos Espec em compostos endógenos no PB, a partir da classificação de Frederiksen (1975) – a “especificação do ambiente”. Com essa abordagem combinada, pretende-se contribuir para o estudo da formação de compostos nominais no PB. A fim de justificar a relevância teórica e a adequação metodológica do estudo, segue sua fundamentação lingüística e psicolingüística.

Fundamentação

Formação de compostos nominais

Do ponto de vista lingüístico, a formação de compostos novos tem sido atribuída a um dos dois níveis da língua: nível lexical e pós-lexical (sintaxe). No nível lexical, uma hipótese de como se formam os compostos é oferecida por Allen (1980), que se apresenta a seguir,¹ segundo citada em Scalise (1987, p.106-15) e adequada à estrutura dos compostos do PB.²

Para Allen, as regras que possibilitam a formação de compostos não são oracionais. São RFP's, mais especificamente, regras de concatenação linear apresentando a seguinte forma:

$$[]_X, []_Y \rightarrow [[]_X Q []_Y]_Z$$

onde Q é o limite interno do composto, impedindo a inserção de um outro termo qualquer entre X e Y. Os compostos estão submetidos a duas condições: a “Condição da variável A”³ e a “Condição Ê UM”.

- Condição da variável A:

prediz que o conteúdo semântico completo do primeiro constituinte pode preencher qualquer das matrizes de traços disponíveis na hierarquia de traços correspondentes ao segundo constituinte, sempre que a matriz de traços a ser pre-

2 A proposta de Allen (1980) foi adequada, principalmente no que tange à ordem dos elementos funcionais nos compostos, que, em inglês, é Espec + Núcleo e no PB, Núcleo + Espec.

3 No original, “variable R”; segundo nota do tradutor em Scalise (1987), R está para “range”. alcance, donde opta-se aqui pelo termo “variável A”.

- Derivação: os compostos pós-lexicais permitem derivação muito restrita (ex.: *roupagem de banho; *proto-roupa de banho; ex-mestre de obras; finzinho de semana⁵).
- Concordância: não há concordância em gênero ou número entre os constituintes do composto, em virtude da estrutura sintática apresentada acima.

Embora Lee não o explicita, pode-se especular que a forma dessa RNM seria a mesma que propôs Dillinger (1993), utilizando a especificação de ambiente, para as relações permitidas para o SN em SPs (tal que P = "com"; "a"), na forma de regras de projeção sintática.

Regra 1 com SN -> ATB: / SN designa atributo

O "com" será interpretado como ATB: quando o núcleo do SN designar um atributo. (...)

(i) Um homem com iniciativa.

(ibidem, p.43 e 45, nota *)

Dillinger utiliza como metodologia de sistematização a "especificação de ambiente", ou a especificação dos dados antes e depois da preposição "com", baseado na classificação semântica de Frederiksen (1975). Note-se que, do ponto de vista lingüístico, a especificação de ambiente parece relevante como metodologia de investigação da formação de compostos novos, pois tanto Allen quanto Dillinger supõem critérios semânticos de seleção de Espec em função do núcleo.

Compreensão de compostos

Também do ponto de vista da realidade psicológica do modelo de linguagem, há razões que tornam a especificação de ambiente uma estratégia atraente para o estudo do comportamento dos falantes. Os falantes percebem linearmente os estímulos verbais (visuais ou auditivos), de modo que seu léxico é primeiramente ativado pelo núcleo e logo

5 Nesse exemplo, seguindo Rocha (1998), admite-se grau como processo de derivação; sob o rótulo de "formação de diminutivo", Lee (1995, p.64) apresenta exemplos semelhantes, onde o núcleo mantém seu estatuto de palavra independente.

em seguida pelo Espec. Dados experimentais sugerem que a velocidade de percepção é maior do que a de ativação do léxico (para revisões recentes, cf. Balota, 1994; Schendan, Ganis & Kutas, 1998), de modo que a ativação do Espec se sobrepõe à do núcleo, incrementando a ativação dos traços em comum e inibindo a de traços distintivos. Chama-se a esse fenômeno “ativação interativa” (Coolen, Jaarsveld & Schreuder, 1993).

Quanto à dimensão sintática dos preposicionados, sabe-se também que os falantes utilizam informações sintáticas para construir proposições, que são um conjunto específico de relações de um termo em função de outro. Pode-se dizer, portanto, que o Espec é selecionado em função do núcleo. Do ponto de vista psicológico, pode-se dizer que as regras de composições seriam um conjunto de funções que explicitam um conjunto de Espec possíveis para um dado conjunto de núcleos, tal como previsto, em níveis distintos, por Allen (1980) e Dillinger (1993). Para o estudo da formação de compostos, portanto, a especificação do ambiente é o primeiro passo, ainda que não seja condição suficiente.

Métodos

Este é um estudo exploratório que visa investigar a formação de compostos endógenos novos. A questão a ser investigada aqui pode ser formulada especificamente como: que tipo de N ocorre no Espec de endógenos em função do núcleo?

Utilizaram-se dois métodos: especificação de ambiente e teste de preenchimento. Para tanto, selecionaram-se 62 compostos endógenos (ver Anexo), classificando-se núcleo e Espec. A partir da classificação desse *corpus* restrito, foi possível estimar o comportamento que se pode esperar dos falantes do PB. Essa estimativa serviu para controlar a produção de compostos no teste de preenchimento. Apresenta-se a seguir cada método em detalhe.

Especificação de ambiente

Neste estudo, a especificação do ambiente foi feita com base nas categorias para objetos, ações e relações propostas por Frederiksen

(1975). Como mostramos, tanto lingüística quanto psicolingüística-mente, há fortes indícios de que o Espec seja selecionado em função do núcleo; daí o foco de investigação recair sobre as relações que se estabelecem entre Espec e núcleo.

Frederiksen (1975) utiliza três níveis categóricos básicos que se superpõem seletivamente. O primeiro nível inclui as classes de "objetos" e "ações". "Objetos são definidos como coisas que ocupam espaço ... Ações são definidas como coisas que ocupam uma posição ou intervalo de tempo e envolvem mudança" (p.378). Traços, tais como localização, tempo, resultado e mudança, são tomados como primitivos. Cada uma dessas classes subdivide-se em subclasses, em função de envolver ou não mudança e de apresentar ou não conteúdo semântico. A Tabela 1 apresenta esquematicamente as classes e subclasses definidas.⁶

Tabela 1 – Classes e subclasses de ações e objetos (cf. Frederiksen, 1975, p.378-97)

1. Ações								
Processos						Ações físicas resultativas		
cognitivos		físicos				cognitivas	físicas	
simples	relativos	cognitivos		não-cognitivos			com resultado simbólico	com resultado não-simbólico
		simples	relativos	simples	relativos			
2. objetos								
estáticos				processivos				
simbólicos		não-simbólicos		animados	inanimados			
					simbólicos	não-simbólicos		

As relações "conectam objetos e ações a outros conceitos, definindo assim 'estados' e 'eventos'" (p.382). Os tipos de relações utilizados para classificação nesse estudo parecem suficientemente claros a partir da Tabela 2, que apresenta as definições utilizadas e exemplifica a classificação aplicada nesse estudo.

⁶ Extrapola o âmbito deste trabalho uma descrição exaustiva, quiçá uma discussão da taxonomia apresentada em Frederiksen (1975); ao leitor interessado, cabe confrontar Dillinger (1993, cap.V) e referências.

continuação

Espec	Núcleo						Frequências	
	AFN	OEN	OES	OPA	PFC	PFN	Absoluta	Relativa
LOC.....	-	4	-	1	-	-	5	0,081
OBJ.....	-	-	1	-	1	-	2	0,03
OBV.....	1	18	6	2	1	-	28	0,452
PRT.....	-	5	-	-	-	-	5	0,081
TEMA.....	-	-	2	1	1	-	4	0,065
Freq. abs.	1	38	10	6	6	1	62	1
Freq. rel.	0,016	0,613	0,161	0,097	0,097	0,016	1	

Como se pode ver, "objeto estático não-simbólico" (OEN) foi o núcleo mais freqüente (61,3% dos compostos); os especificadores "objeto" (OBV) foram mais freqüentes (45,2%); OENs apresentaram-se especificados principalmente por OBVs (47,4%). No entanto, a interdependência entre classe do núcleo (OEN ou não-OEN) e classe do Espec (OBV ou não-OBV) mostrou-se estatisticamente insignificante ($\chi^2 = 0,193$).

A partir dessas informações, supôs-se que um ambiente natural para preposicionados seria o par OEN-OBV. Isto dirigiu a construção do teste de preenchimento para levantamento de dados, conforme se expõe a seguir.

Teste de preenchimento

Material

O teste de preenchimento foi desenvolvido originalmente para levantamento de dados de aceitabilidade de estruturas sintáticas (Trueswell et al., 1993; Garnsey et al., 1997). É de aplicação extremamente simples, ao mesmo tempo que permite o levantamento de um número razoavelmente grande de dados. Nesse estudo, substituíram-se os itens sintáticos por compostos endógenos.

Trata-se de um questionário de resposta aberta, apresentado na forma de uma lista com duas colunas, contendo a da esquerda as seguintes instruções:

Complete com um substantivo os itens ao lado, de modo que:

1. O item forme uma unidade de sentido;
2. O item formado seja UM TIPO disso que está escrito.

Exemplo: CAMPO DE FUTEBOL é um tipo de campo.

Obrigado por colaborar.

Na coluna da direita, o questionário apresentado continha 24 itens do tipo "[OEN] de _____" (doravante, estímulos), a ser preenchido conforme as instruções.

Estímulos

Para controlar a variabilidade dos resultados em função do núcleo, todos os 24 estímulos apresentados foram do tipo OEN. Destes, 14 itens podem ser classificados como de "alta expectativa" (AE), seguindo Coolen et al. (1991). Tais itens participam como núcleo em endógenos lexicalizados, sendo possível esperar como resposta a recuperação de itens do léxico. Os estímulos AE foram todos escolhidos dentre compostos especificados com OBV, a fim de direcionar a expectativa. Em contraste, 10 itens são de "baixa expectativa" (BE), pois não participam de endógenos lexicalizados. Nesses casos, ao preencher os itens, os sujeitos estariam optando por alguma característica que seria especificada no composto formado. Assim, AE serve para distrair a atenção do participante, oferecendo-lhe um ambiente natural de ocorrência de endógenos. A produção em resposta a BE é o foco principal do estudo. A Tabela 4 apresenta os itens de estímulo, exemplificando compostos de que participam os itens AE e possíveis relações no caso de itens BE.

Tabela 4 – Estímulos de alta e baixa expectativa

AE		BE	
Itens	Compostos	Itens	Relação
colher	colher de pau	avental	LOC/INIT
xicara	xícara de chá	estante	OBV/INIT
carro	carro de corrida	barco	OBV/INIT
óculos	óculos de sol	mangueira	OBV/INIT
pasta	pasta de dentes	frasco	LOC1/INIT
papel	papel de presente	mala	OBV/INIT
cadeira	cadeira de balanço	lata	LOC1/INIT
caixa	caixa de fósforos	disco	OBV/INIT
posto	posto de gasolina	copo	LOC1/INIT
casa	casa de boneca	palco	LOC/OBV
câmara	câmara de gás		
sala	sala de aula		
roupa	roupa de cama		
cama	cama de casal		

Participantes

Participaram 40 alunos de duas turmas dos cursos de Psicologia e Letras da UFMG. Para isolar efeitos de ordem – em que a ordem dos estímulos afeta sistematicamente os resultados –, os mesmos estímulos foram apresentados em quatro ordens distintas: listas A e B apresentavam estímulos em ordem aleatória e listas C e D, em ordem pseudo-aleatória, em que aos itens AE intercalavam-se os BE. Cada lista foi respondida por dez sujeitos, durante o horário normal da aula, dentro de sala.

Resultados e discussão

Conforme se disse, o objetivo específico do teste foi o de verificar quais Espec seriam priorizados na produção de endógenos, em resposta a BE. Os compostos apresentados foram classificados de acordo com a Tabela 2 (ver anexo). Excluíram-se posteriormente todos os compostos formados a partir do estímulo "palco", porque as distinções pareceram não ser suficientemente claras. Excluíram-se ainda aqueles que não se coadunavam com as instruções (p. ex.: *roupa de dormir*). As ocorrências analisadas representam 93,5% das ocorrências possíveis (Tabela 5).

Tabela 5 – Resultados do teste de preenchimento (n = 898)

Espec	Núcleo			
	Tipo		Frequência	
	AE	BE	Absoluta	Relativa
OBV.....	359	150	509	0,57
PRT.....	114	146	260	0,29
OBJ2.....	12	48	60	0,07
Outros.....	59	10	69	0,08
TOTAL	544	354	898	1

A interdependência entre a classe dos núcleos (AE ou BE) e a classe dos Espec, tal como apresentado na Tabela 5, revelou-se extremamente significativa (χ^2 , $p = 6,88724E-24$). A indução ao par AE-OBV também se revelou significativa, na medida em que foi grande a diferença no teste de duas amostras para proporções de Espec OBV dentre

AE e BE ($z = 5,853$).⁷ Assim, a produção em resposta a BE pode ser resumida da seguinte maneira.⁸

a) Ao contrário dos Espec de AE, a proporção de OBV não difere significativamente daquela encontrada no *corpus* ($z = 1,037$);

b) OBV também foi o Espec mais realizado (44,6%);

c) OBV, PRT e OBJ2 respondem por 98,3% dos casos;

Comparações diretas, entre as produções para AE e BE ou dessas com o *corpus*, devem ser tomadas *cum granum salis*. O *corpus* é bastante restrito para que se possa generalizar a ocorrência de Espec em função de núcleo. Além disso, o teste foi construído de tal modo que apenas permite focalizar em BE, uma vez que os núcleos AE foram escolhidos dentre compostos existentes e permitem recuperação de itens da memória, sem necessidade de aplicação de regras (cf. item Estímulos). O item (a) mostra claramente que a grande ocorrência de OBV para AE deve ser creditada ao viés do teste.

O fato de OBV ter sido muito realizado em BE deve ser considerado, por outro lado, como um comportamento genuíno, não derivado da influência de AE. A ativação na mente do falante dá-se na ordem da leitura, isto é, do núcleo para o Espec. Além disso, mesmo para AE, a produção de OBV não foi categórica (64,5%), ao contrário do que os itens permitiam. É de se esperar, portanto, que uma investigação mais abrangente, em *corpus* falado ou escrito, indique uma predominância de Espec OBV para OEN.

Ao contrário da opinião generalizada, a formação de compostos parece bastante sistemática, tal como representada neste estudo pelos compostos formados em resposta aos estímulos BE (item c). Os resultados aqui apresentados podem ser resumidos na seguinte regra:

$$[N]_x \text{ de } [N]_y, x = \text{OEN} \rightarrow [[\text{OEN}]_x \text{ de } \{\text{PRT; OBV; OBJ2}\}_y]_z$$

Na classificação de Frederiksen, OBJ2 refere-se a “objeto não relacionado”. Em *copo de água*, água não está relacionada ao copo como parte específica (PRT) ou o material de que é feito (INIT). Porém, parece que, para a classificação de endógenos, a distinção *a priori* de OBJ2 e OBV indica uma diferença de uso e não de estrutura. Lembre-se da dis-

7 Compare-se esse valor com o valor de Z para uma probabilidade de erro de 1% ($z = 2,58$) e 5% ($z = 1,96$).

8 A categoria “outros” inclui os Espec LOC, AGT, ATB e INIT, cuja análise, isoladamente, não se revelou importante.

tingão entre restrições estruturais e contextuais, tal como se aplicam aos fenômenos de derivação. Tome-se o exemplo da regra

N → N -eiro.

Essa regra especifica que um nome [-humano] dará origem a um agentivo relacionado. É o que ocorre em vaca – vaqueiro. Poderão ocorrer outros produtos, desde que respeitadas as definições da regra. Sabe-se que daí à existência real do produto, uma série de outras exigências do uso da língua devem ser respeitadas, tais como a existência de outro item de mesma função (que provoca um bloqueio) e a necessidade mesma do item para a comunicação. Essa definição, que remonta ao trabalho de Aronoff (cf. Rocha, 1998), distingue palavras possíveis de palavras reais. Nada impediria, portanto, que ao se defrontar pela primeira vez com um produto da regra, o falante seja capaz de interpretá-lo adequadamente, como mostram os exemplos abaixo:

- a. A Gaviões da Fiel aprovou o filme "Boleiros", de Ugo Giorgetti.
- b. – Já 'tá quase na hora do aniversário e o bolo ainda não chegou...
- b'. – Liga lá p'ra casa da boleira.

Em (a), "boleiros" é facilmente compreendido como produto de BOLA – boleiro (refere-se aos torcedores fanáticos, tema do filme). Em (b), a base não será compreendida como bola, mas BOLO. O fato de não ser lexicalizada não significa que a formação em (b') não seja possível e sistemática, ao menos como formação temporária.

Do mesmo modo, seria adequado estipular dois níveis de análise para os endógenos. No nível da existência real dos produtos, algumas contingências não estruturais levariam à gramaticalização dos itens. No nível das regras de composição, ter-se-ia uma especificação das condições estruturais em que se poderia formar uma unidade, que aqui interessou de perto. Com isso, teríamos uma análise diferente dos dados. Ao fazer a paráfrase, viu-se, por exemplo, que *frasco de perfume* pode ser compreendido como "frasco para perfume", pois pode ser reconhecido como tal, mesmo que não contenha o líquido. Sejam os endógenos já cristalizados na língua:

- c. frasco de perfume
- d. copo de água
- e. carro de bombeiro

Nesta análise, classificaram-se (c, e) como OBV, por causa do exposto acima, sobre as características físicas de (c). Seguindo Allen (1980),

"copo de água" não É UM tipo de copo, é um NP simples, um copo contendo água. Mas o que o distingue de (c), classificado como OBV? Apenas e tão-somente o fato de não existir realmente como designador de um referente único. Desse modo, os dados deste estudo seriam ainda mais categóricos quanto à produção de endógenos. Para verificar isto, reanalisaram-se os dados, consolidando as classes OBV e OBJ2 (Tabela 6).

Tabela 6 – Resultados reagrupados

Espec	Núcleo			
	Tipo		Frequência	
	AE	BE	Absoluta	Relativa
OBV+OBJ2.....	371	198	569	0,63
PRT.....	114	146	260	0,29
Outros.....	59	10	69	0,08
TOTAL	544	354	898	1

O teste de independência sobre os dados reagrupados revelou-se ainda bastante significativo ($p = 5,54766E-12$), porém menos do que o teste sobre os dados da tabela 5.

A investigação do comportamento dos falantes do PB, com relação a uma classe reunida OBJ2 e OBV, poderá ser objeto de estudos futuros, em que se manipule diferentemente a construção do teste. Caberá ainda verificar a relação entre OBJ2/OBV e OEN, do ponto de vista de sua compreensão. Onde o traço de continente é a dimensão especificada, é possível observar uma semelhança interessante entre os itens apresentados e alguns compostos exógenos de estrutura semelhante. Trata-se de compostos como *colher de açúcar*, utilizado para expressar UMA QUANTIDADE determinada, não um objeto. Nesse caso, o mecanismo de associação metonímica "continente pelo conteúdo" pode ser o que permite especificar não a colher (OEN), mas a quantidade que cabe na colher (ATB). O mesmo mecanismo pode ter permitido, no teste, o aparecimento de "lata de leite" como um tipo de lata: os sujeitos também selecionaram o traço "continente". Nesse caso, é possível sugerir uma regra de composição, capaz de generalizar uma tendência de formação de preposicionados novos em PB:

$$[OEN(+continente)]x \text{ de } [OEN]y \rightarrow [[OEN]x [OBJ2+OBV]]z$$

Essa regra generaliza a maioria dos dados levantados no teste, como mostra a Tabela 6. A demonstração psicológica dessa regra, no

entanto, é tarefa de um teste de compreensão de compostos novos isolados (conforme realizado por Coolen et al., 1991).

Agradecimentos

Agradecemos a Mike Dillinger, João H. R. Tótar, Luiz Carlos Rocha e, em especial, a Vitor G. Haase, Carlos A. Gohn e alunos (jan. 1998). A elaboração final contou com apoio da Fapemig (BDT n.20634/98).

ROTHER-NEVES, R. Nominal compound formation in Brazilian Portuguese: a psycholinguistic investigation. *Alfa (São Paulo)*, v.43, p.185-202, 1999.

- *ABSTRACT: This study investigated which specifiers occur more often in the making of nominal compounds with structure N-de-N whose nucleus is a "static non-symbolic object". Data from native Brazilian Portuguese speakers (n = 40) were collected with a completion test made after a semantic analysis of 62 compounds, in accordance to Frederiksen (1975). Highly expected items served as distractors. This study focused on items with low expectance, which had almost exclusively (98,3%; n = 898) specifiers acting as "part", "goal" and "unrelated object". The distinction between the last two is discussed in the light of the distinction between possible and real existing words. A hypothesis is presented, that the semantic feature "continent" is select in the formation of nominal compounds with the same nucleus, which is to be further investigated.*
- *KEYWORDS: Word formation; nominal compounds; psycholinguistics.*

Referências bibliográficas

- ALLEN, M. R. Semantic and phonological consequences of boundaries: a morphological analysis of compounds. In: ARONOFF, M., KEAN, M. L. *Juncture*: Amna Libri, 1980. Apud: SCALISE, S. *Morfologia generativa*. Trad. José Pazó. Madrid: Alianza, 1987. p.106-15.
- ARONOFF, M. *Word formation in generative grammar*. Cambridge, MA: MIT, 1976.
- BADECKER, W., ZANUTTINI, R., MIOZZO, M. The two-stage model of lexical retrieval: evidence from a case of anomia with selective preservation of grammatical gender. *Cognition (Amsterdam)*, v.57, n.2, p.193-216, 1995.
- BALOTA, D. Visual word recognition: the journey from features to meaning. In: GERNSBACHER, M. A. (Org.) *Handbook of psycholinguistics*. San Diego: Academic Press, 1994.

- BASÍLIO, M. *Teoria lexical*. 4.ed. São Paulo: Ática, 1995. (Série Princípios, 88).
- CARAMAZZA, A. Data, statistics and theory: a comment on Bates, McDonald, MacWhinney, and Applebaum's "A maximum likelihood procedure for the analysis of group and individual data in aphasia research". *Brain and Language (San Diego)*, v.41, p.43-51, 1991.
- COOLEN, R., JAARSVELD, H. J. van, SCHREUDER, R. The interpretation of isolated novel nominal compounds. *Journal of Memory and Cognition*, v.19, n.4, p.341-52, 1991.
- _____. Processing novel compounds: evidence for interactive meaning activation of ambiguous nouns. *Journal of Memory and Cognition*, v.21, n.2, p.235-46, 1993.
- DILLINGER, M. *Notas sobre semântica*. Belo Horizonte: Departamento de Linguística da FALE/UFMG, 1993. (Mimeogr.).
- DI SCIULLO, A. M., WILLIAMS, E. *On the definition of word*. Cambridge: MIT Press, 1987.
- FREDERIKSEN, C. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology*, v.7, p.371-458, 1975.
- GARNSEY, S. M. et al. Contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, v.37, p.58-93, 1997.
- LEE, S.-H. *Morfologia e fonologia lexical do Português do Brasil*. Campinas, 1995. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem, Universidade de Campinas.
- MICELI, G. Morphological errors and the representation of morphology in the lexical-semantic system. *Phil. Trans. Royal Soc. London B*, v.346, p.79-88, 1994.
- NADEAU, S. E. Impaired grammar with normal fluency and phonology. Implications for Broca's aphasia. *Brain (Oxford)*, v.111, p.1111-37, 1988.
- RAPP, B. C. CARAMAZZA, A. Disorders of lexical processing and the lexicon. In: GAZZANIGA, M. S. (Org.) *The Cognitive Neurosciences*. Cambridge (MA): MIT Press, 1995. p.901-14.
- ROCHA, L. C. de A. *Estruturas morfológicas do português*. Belo Horizonte: Editora da UFMG, 1998.
- SANDMANN, A. J. *Morfologia geral*. 2.ed. São Paulo: Contexto, 1993.
- SCHENDAN, H. E., GANIS, G., KUTAS, M. Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology*, v.35, n.3, p.240-51, 1998.
- TRUESWELL, J. C., TANENHAUS, M. K., KELLO, C. Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psy.: Lang. Mem. Cognition*, v.19, p.528-53, 1993.
- TYLER, L. K. Processing distinctions between stems and affixes: evidence from a non-fluent aphasic patient. *Cognition*, v.36, n.2, p.129-53, Aug. 1990.
- _____. Use of derivational morphology during reading. *Cognition*, v.36, n.1, p.17-34, July 1990.

de “preposicionados”, e, dentre esses, nos chamados “compostos endógenos” (Sandmann, 1993, p.60-9). No PB, compostos preposicionados apresentam a seguinte estrutura funcional (cf. Basílio, 1995, p.29):

$$[[\text{Núcleo}]_X [\text{Especificador}]_Y]_Z$$

Nos compostos endógenos, Z é um tipo específico de X. Em *óculos de sol*, compreende-se tratar de um tipo específico de *óculos*. O mesmo não ocorre com compostos exógenos, em que o significado do composto é metafórico ou metonímico, não podendo ser derivado de seu núcleo: *lua-de-mel* não é um tipo de lua. O problema que se coloca, então, é: no composto endógeno, quais serão as condições de generalização de Y em função de X, tal que se possa descrever a formação de Z.

Muitos autores (incluindo Basílio e Sandmann) limitam-se a sistematizar compostos existentes, o que contribui para a opinião generalizada de que a composição envolve processos idiossincráticos. Essa opinião, no entanto, não pode adequar-se a duas noções bastante claras. Em primeiro lugar, quaisquer N separados por “de” não formam um preposicionado (*borboleta de afasia). Em segundo lugar, dado um N determinado na posição de núcleo, nem todo N pode ocupar a de especificador (Espec). Assim, o Espec de “cor de _” muito provavelmente não será um evento, ainda que algum evento esteja implícito (*cor-de-fogo*).

As regras de formação de palavras (RFP's) são um construto teórico proposto por Aronoff (1976), para dar conta das regularidades morfológicas. Referem-se ao que permite aos falantes de uma língua criar palavras novas, a partir de itens já existentes. Uma grande contribuição ao seu estudo vem sendo oferecida pela pesquisa psicológica da linguagem, capaz de fornecer evidências de processamento distinto para bases e afixos (Tyler et al., 1990) e de uma organização morfológica do léxico (Tyler & Nagy, 1990), com armazenamento independente de informações semânticas e traços morfológicos (Badecker, Zanuttini, Miozzo, 1995), bem como da distinção dos processamentos morfológico e sintático (Nadeau, 1988). Essas evidências são coletadas por meio de estudos tanto com sujeitos normais quanto com pacientes que sofrem de algum déficit seletivo de linguagem (Caramazza, 1991; Miceli, 1994; Rapp & Caramazza, 1995). Tais estudos servem para detalhar mais precisamente os resultados da investigação lingüística com relação a sua validade psicológica.

No presente estudo, utilizou-se uma adaptação do procedimento psicolingüístico conhecido como “teste de preenchimento” (*completion test*, cf. Trueswell, Tanenhaus, Kello, 1993; Garnsey et al., 1997) em

enchida corresponda a algum dos traços do primeiro elemento (Allen, 1980, p.93, cit. in Scalise, 1987, p.107).

Assim, em um composto qualquer, seu significado resulta da seleção de um traço específico do núcleo em função de traços específicos do Espec.

- Condição da variável É UM: estipula que Z É UM X em um composto $[[]_X []_Y]_Z$ no PB. A formulação de Allen aplica-se tanto à sintaxe quanto à semântica, isto é, o composto resultante será da mesma categoria sintática que seu núcleo e “se estabelece um subconjunto de relações semânticas entre o composto Z” e o núcleo (Allen, 1980, p.11, apud Scalise, 1987, p.109).

É interessante notar que Scalise explicitamente vincula a previsibilidade dos compostos às condições apresentadas. Segundo ele, apenas seriam compostos regulares os que obedecessem a essas condições, enquanto os outros seriam formações lexicais, “pelo que podemos considerá-los compostos ‘lexicalizados’” (p.110).

Para Lee (1995, cap.3), os endógenos não se formam no léxico, mas a partir de regras sintáticas. Lee parte dos pressupostos da Morfologia Lexical e distingue dois tipos de compostos no PB: compostos lexicais e pós-lexicais.

[C]ompostos lexicais são “compostos verdadeiros”, que funcionam como unidade independente nas operações morfológicas. (Lee, 1995, p.53)

Os compostos pós-lexicais levam esse nome por serem formados no nível pós-lexical (sintático), por meio de uma “regra não-morfológica de criação de palavra” (RNM; ingl.: nonmorphological word-creating rule), termo emprestado a Di Sciullo & Williams (1987). Os preposicionados seriam do tipo composto pós-lexical, dados os seguintes fatos sintáticos,⁴ além da ordem dos elementos funcionais (já apresentada anteriormente):

- Formação de plural: o morfema de plural ocorre ENTRE os constituintes do composto, dada a seguinte estrutura sintática:

[NP[N *roupas*] [PP[P de] [NP *banho*]]]

4 Nem o trabalho de Lee versa especificamente sobre preposicionados, nem o âmbito do presente estudo permitiu descer aos detalhes de sua abordagem. Os exemplos a seguir foram escolhidos pelo autor, com base na discussão oferecida por Lee (1995, p.62-7).

Tabela 2 – Classificação de Núcleo e Especificador dos itens do *corpus*

CLASSE	EXEMPLO*
Objetos	
OEN objeto estático não-simbólico	mármore de carrara
OES objeto estático simbólico	aviso de recebimento
OPA objeto processivo animado	ama de leite
Ações	
AFN ação física (com resultado) não-simbólico	viagem de estudos
PFC processo físico cognitivo	defesa de tese
PFN processo físico não-cognitivo	mal de parkinson
Relações	
AGT agente	palavra de escoteiro
ATB atributo	palavra de honra
CAU causa	dores de parto
EVT evento	calendário de eventos
INIT material ou procedência	palha de aço
LOC localização	mesinha de cabeceira
OBJ objeto afetado	coleta de lixo
OBJ2 objeto não-relacionado	lata de leite
OBV objetivo	carro de corrida
PRT parte componente	cadeira de rodas
TEMA tema	aviso de recebimento

* O termo em negrito exemplifica a classe.

Para fins de análise, parafrasearam-se os compostos para melhor discernimento da relação envolvida. Assim, palha de aço pode ser parafraseada como “feita de aço”, relação INIT; casa de praia, paráfrase “na praia”, relação LOC. Núcleos e Espec foram assim tabulados para análise de frequência (Tabela 3).

Tabela 3 – Ocorrência de Espec em função de núcleo no *corpus* (n = 62)

Espec	Núcleo						Frequências	
	AFN	OEN	OES	OPA	PFC	PFN	Absoluta	Relativa
AGT.....	-	-	-	-	1	-	1	0,016
ATB.....	-	6	-	1	2	-	9	0,15
CAU.....	-	-	-	-	-	1	1	0,016
EVT.....	-	-	1	-	-	-	1	0,016
INIT.....	-	5	-	1	-	-	6	0,097

Anexo 1 – Classificação dos endógenos do corpus

Núcleo	Composto	Espec.	Núcleo	Composto	Espec.	Núcleo	Composto	Espec.	Núcleo	Composto	Espec.
AFN	viagem de estudos	OBV	OEN	pasta de dente	OBV	OES	fundo de garantia	OBV	OES		OBV
OEN	feijão-de-corda	ATB	OEN	papel de parede	OBV	OES	vara de pesca	OBV	OES		OBV
OEN	cachorro de madame	ATB	OEN	papel de presente	OBV	OES	convite de casamento	OBV	OES		OBV
OEN	papel de seda	ATB	OEN	cadeira de balanço	OBV	OES	texto de partida	OBV	OES		OBV
OEN	camisa-de-força	ATB	OEN	caixa de sapato	OBV	OES	cartão de crédito	OBV	OES		OBV
OEN	fogo de artifício	ATB	OEN	ferro de engomar	OBV	OES	aviso de chegada	TEMA	OES		TEMA
OEN	pimenta-de-cheiro	ATB	OEN	carro de bombeiro	OBV	OES	aviso de recebimento	TEMA	OES		TEMA
OEN	pó-de-arroz	INIT	OEN	posto de gasolina	OBV	OPA	banda de rock	ATB	OPA		ATB
OEN	palha de aço	INIT	OEN	casa de detenção	OBV	OPA	galinha d'angola	INIT	OPA		INIT
OEN	mármore de Carrara	INIT	OEN	creme de barbear	OBV	OPA	rato de esgoto	LOC	OPA		LOC
OEN	canela-da-índia	INIT	OEN	câmara de ar	OBV	OPA	ama-de-leite	OBV	OPA		OBV
OEN	vinho do porto	INIT	OEN	sala de jantar	OBV	OPA	mestre-de-obras	OBV	OPA		OBV
OEN	mesa de bar	LOC	OEN	taça de champagne	OBV	OPA	crítico de cinema	TEMA	OPA		TEMA
OEN	mesa de escritório	LOC	OEN	dropes de anis	PRT	PFC	palavra de escoteiro	ACT	PFC		ACT
OEN	livro de cabeceira	LOC	OEN	sopa de letrinhas	PRT	PFC	conversa de botequim	ATB	PFC		ATB
OEN	mesa de cabeceira	LOC	OEN	cadeira de rodas	PRT	PFC	palavra de honra	ATB	PFC		ATB
OEN	colher de sopa	OBV	OEN	giz de cera	PRT	PFC	coleta de lixo	OBJ	PFC		OBJ
OEN	xícara de café	OBV	OEN	pão-de-ló	PRT	PFC	ação de graças	OBV	PFC		OBV
OEN	carro de corrida	OBV	OES	calendário de atividades	EVT	PFC	defesa de tese	TEMA	PFC		TEMA
OEN	óculos de sol	OBV	OES	exame de sangue	OBJ	PFN	dores de parto	CAU	PFN		CAU
OEN	pasta de trabalho	OBV	OES	cartão de embarque	OBV						

Anexo 2 – Ocorrências de Espec por item no teste

	ESPEC							Total
	PRT	OBV	OBJ2	LOC	AGT	ATB	INIT	
AE	114	359	12	12	2	29	16	544
colher	20	17	3	-	-	-	-	40
xícara	5	34	1	-	-	-	-	40
carro	3	18	-	-	-	5	13	39
óculos	3	25	-	-	-	9	1	38
pasta	12	28	-	-	-	-	-	40
papel	1	24	-	-	-	15	-	40
cadeira	15	22	-	-	-	-	2	39
caixa	13	17	6	1	1	-	-	38
posto	-	40	-	-	-	-	-	40
casa	17	14	-	7	1	-	-	39
câmara	4	34	-	1	-	-	-	39
sala	-	40	-	-	-	-	-	40
roupa	5	27	-	3	-	-	-	35
cama	16	19	2	-	-	-	-	37
BE	146	150	48	0	0	6	4	354
barco	23	12	-	-	-	-	4	39
avental	17	23	-	-	-	-	-	40
estante	18	22	-	-	-	-	-	40
mangueira	17	23	-	-	-	-	-	40
frasco	2	28	8	-	-	-	-	38
mala	9	24	7	-	-	-	-	40
lata	4	18	17	-	-	-	-	39
disco	32	-	-	-	-	6	-	38
copo	24	-	16	-	-	-	-	40
TOTAL	260	509	60	12	2	35	20	898