4
Composicionalidade com base em córpus

> *Queremos estabelecer uma ordem no nosso conhecimento da linguagem: uma ordem para uma finalidade determinada; uma ordem dentre as muitas possíveis, não a ordem.*
>
> Wittgenstein, *Investigações Filosóficas*

Neste capítulo iremos focar a última etapa de implementação computacional do nosso estudo para aplicação de uma medida de composicionalidade semântica em relação às CMs detectadas no capítulo 3. Trata-se de uma medida de similaridade entre os microcontextos (parágrafos do córpus) em que as CMs ocorrem (cf. Garrão, Oliveira, Freitas & Dias, 2006). É importante adiantar que nosso critério de aferição do grau de transparência semântica se baseia em uma técnica utilizada no domínio computacional de Recuperação de Informação (RI), em detrimento de uma aferição intuitiva, com base nos testes já criticados na seção 2.3.1.

Na verdade, muito se especula sobre a importância da aplicação de teorias semânticas já existentes na lingüística, como aquelas apresentadas e questionadas no capítulo 2, para fins de PLN. Contrariamente, pretendemos demonstrar neste capítulo a importância de PLN e do córpus para avaliar semanticamente as CMs em questão. Através de uma medida de similaridade entre os microcontextos em que uma dada CM aparece e os microcontextos em que detectamos apenas o SN que compõe a CM, pretendemos avaliar empiricamente o que é dito ser "transparência e opacidade semântica".

Nossa proposta é aferir o grau de transparência/opacidade semântica de uma CM pelos contrastes entre os contextos de uso da CM propriamente dita (como por exemplo, *fazer campanha*) e os contextos de uso do SN que compõe a CM (*campanha*) em detrimento de uma avaliação semântica apriorística dos itens que compõem a CM. Em outras palavras, é o que está fora da CM que vai determinar o seu grau de composicionalidade semântica, como demonstraremos na seção 4.2.

Conforme apresentados na seção 2.3.1, os critérios tradicionalmente utilizados para caracterizar um segmento lingüístico como uma CM são:

i) não-composicionalidade – o significado do todo não corresponde à soma das partes, como o exemplo já criticado no capítulo 2, *bater as botas*. (Guenthner e Blanco, 2004; Neves, 1999, entre outros). Um dos problemas desta definição é que outros autores argumentam também que é possível caracterizar algumas CMs pela possibilidade de seus componentes contribuírem para a semântica do composto (formando uma *colocação*) em oposição ao conceito de *expressões idiomáticas*, como por exemplo, a distinção entre a colocação *pagar as contas* e a expressão idiomática *pagar mico* (Cruse, 1986)**.** Segundo o autor, a composicionalidade seria o melhor critério para distinguir uma colocação de uma expressão idiomática, embora admita que a distinção seja difícil em muitos casos.

ii) não-substituição ou arbitrariedade - não é possível substituir palavras que compõem uma CM mantendo a integridade da expressão, ainda que a palavra substituída seja sinônima da original (Tagnin, 1999; Mannning e Schütze, 1999). Este critério pode ser contra-argumentado até pela expressão exaustivamente utilizada para definir a noção de opacidade semântica: *bater as botas – bater a caçuleta*. Tal critério também se fia em uma questão problemática na Semântica: como podemos determinar que uma palavra é sinônima da outra? Pode-se dizer, por exemplo, que, em determinados contextos, *dar uma festa* e *fazer uma festa* são CMs sinônimas; por outro lado, será que uma visão semântica tradicional consideraria *dar* e *fazer* verbos sinônimos?

iii) não-modificação - muitas CMs não podem ser livremente modificadas pela adição de informação lexical ou de transformações gramaticais (Guenthner e Blanco, 2004). Contrariando este último critério, Cruse (1986) afirma que colocações variam quanto ao número de palavras envolvidas, quanto às relações sintáticas entre as palavras, e quanto ao grau de rigidez com que os itens são combinados ("*tomar uma decisão*"/ "*uma decisão foi tomada*"/ "a *tomada de decisões*"). Este terceiro critério fica ainda mais complicado quando a CM não é nominal, como *alto falante* ou *criado mudo,* mas verbal, com várias possibilidades aspectuais: como *receber tratamento, receber o mesmo tratamento; tomar decisão; tomar várias decisões.* Pudemos verificar em Garrão (2001) que até mesmo as CMs consideradas altamente opacas como *bater perna, fazer questão, pagar mico* podem sofrer modificação através de inserção de um marcador aspectual: *bater muita perna, fazer muita questão, pagar o maior mico*.

Tais opiniões contraditórias sobre a detecção de CMs — baseadas naquilo que rotulamos no capítulo 2 como semântica do cálculo e naquilo que Fillmore (1979) rotulou como semântica da inocência —, nos impulsionou a procurar uma via alternativa para caracterizar o grau de composicionalidade de uma CM. Segundo Aranha, Freitas, Dias & Passos (2004), "palavras com significados similares tenderão a ocorrer em contextos similares e palavras polissêmicas tenderão a ocorrer em contextos diferentes".

Tomando como pressuposto essa idéia, consideramos possível fazer avaliações sobre o grau de similaridade entre microcontextos contendo uma CM e microcontextos contendo somente o SN que compõe a CM. A hipótese, portanto, é a de que o grau de transparência semântica da CM é proporcional ao aumento do grau de similaridade entre os parágrafos contendo uma certa CM e os parágrafos contendo somente o SN presente na CM. Em outras palavras, se os parágrafos do córpus que contêm todas as CMs *fazer campanha* forem similares aos parágrafos do córpus que contêm o SN *campanha*, a probabilidade de se encontrar *fazer campanha* nos mesmos microcontextos de *campanha* é muito grande. Isso indicaria que a CM é transparente.

Além de eliminar o risco da aferição intuitiva de composicionalidade semântica das CMs, este recurso se revelou também vocacionado a detectar o grau de polissemia dos SNs quando não pertencem à CM, como também do grau de polissemia das CMs propriamente ditas, como veremos na seção 4.2.

4.1
Passo-a-passo do método

Um dos métodos mais utilizados para medir o grau de similaridade entre documentos é baseado em um *Modelo de Espaço Vetorial* (Baeza-Yates & Ribeiro-Neto, 1999). Esse modelo representa os documentos através de todas as palavras neles contidas. Por meio dele, pode-se estabelecer o grau de similaridade entre os documentos. Cada documento, ou parágrafo (que chamamos aqui de microcontexto) é dividido em uma tabela de freqüência de palavras. As tabelas são chamadas de vetores. Um vocabulário é construído a partir de todos os

microcontextos. Cada microcontexto é representado como um vetor em relação ao vocabulário total dos microcontextos.

Exemplo simplificado:

Documento A:

Um cachorro e um gato.

um	cachorro	e	gato
2	1	1	1

Documento B

Um sapo

um	sapo
1	1

O vocabulário é a soma de todas as palavras utilizadas, isto é, de todos os documentos:

um, cachorro, e, gato, sapo

Portanto:

Documento A:

Um cachorro e um gato.

um	cachorro	e	gato	sapo
2	1	1	1	0

Vetor: (2,1,1,1,0)

Documento B

Um sapo

um	cachorro	e	gato	sapo
1	0	0	0	1

Vetor: (1, 0, 0, 0, 1)

Este exemplo bastante simplificado demonstra como podemos estabelecer entre documentos medidas de similaridades, sendo que, numa aferição real, as

palavras funcionais (*um, e*, no exemplo acima) são descartadas (chamadas de "*stop words"*). Essas medidas são altamente relevantes para aplicações computacionais que lidam com a Teoria da Informação (TI), como sistemas de Recuperação de Informação (como, por exemplo, o *Google*™). Quanto mais palavras os documentos tiverem em comum, mais similares serão entre si, maior a relação entre eles. É essa técnica que facilita a pesquisa do usuário em um sistema de busca; por meio dela, é possível estabelecer o grau de semelhança entre inúmeros documentos disponíveis na rede.

Neste estudo, portanto, nos apropriamos desta mesma idéia de similaridade entre documentos para detectar a relação entre os microcontextos que contenham uma certa CM e os microcontextos que contenham somente o SN fora da CM. Tomemos como exemplo a aferição do grau de similaridade entre os microcontextos de *fazer campanha* e *campanha*: primeiramente, aferimos o grau de similaridade entre todos os parágrafos do córpus CETENFolha que contenham as CMs *fazer campanha* e calculamos a sua similaridade média. Posteriormente, aferimos o grau de similaridade entre todos os parágrafos do córpus CETENFolha que contenham os SNs *campanha* (sem estarem precedidos pelo verbo "fazer" ou por qualquer outro verbo) e calculamos a similaridade média entre eles. Por fim, calculamos a similaridade média entre os parágrafos que contêm a CM e aqueles que contêm o SN.

Quanto maiores forem os números expostos nos vetores de cada microcontexto em relação ao vocabulário total dos microcontextos, maior a relação entre os microcontextos das CMs e dos SNs e, conseqüentemente, maior o grau de composicionalidade semântica da CM. Quanto menores forem esses números, menor a relação entre os microcontextos, e maior o grau de opacidade semântica da CM.

4.2
Aferição do grau de composicionalidade das CMs

Depois da identificação das CMs mais freqüentes do córpus, nós observamos os contrastes entre os microcontextos em que aparecem, através do

Modelo de Espaço Vetorial. As duas expressões cujos graus de similaridade pretendemos aferir (como, por exemplo, *fazer campanha* e *campanha*) são representadas como vetores em um espaço multidimensional. O cosseno entre esses dois vetores indica as palavras que eles têm em comum e, por essa razão, o método pode ser considerado como uma medida de similaridade entre dados.

Para cada CM *w* nós realizamos as seguintes etapas:

i) extração de todos os parágrafos contendo *w* (conjunto *P*1; por exemplo *fazer campanha);*

ii) extração de todos os parágrafos contendo o substantivo em *w* que não ocorre em *P*1 (conjunto *P*2; *campanha);*

iii) indexação de *P*1 e *P*2 no Modelo de Espaço Vetorial;

iv) cálculo das matrizes de similaridades entre parágrafos em *P*1 e obtenção da média dos seus valores;

v) cálculo das matrizes de similaridades entre parágrafos em *P*2 e obtenção da média de seus valores;

vi) cálculo das matrizes de similaridades entre os parágrafos em *P*1 e *P*2 e obtenção da média de seus valores.

Portanto, o aumento das similaridades entre os parágrafos em P1 e P2 é proporcional ao aumento do grau de composicionalidade da CM em questão. Para avaliar tal hipótese, é calculada a similaridade intra-*P*1 (entre todos os microcontextos que contêm *fazer campanha*) , e em seguida intra-*P*2 (entre todos os microcontextos que contêm *campanha*). Finalmente, é avaliada a similaridade entre *P*1 e *P*2. É esta última etapa que nos dará o grau de composicionalidade semântica da CM.

É importante mencionar que escolhemos pelo menos 30 ocorrências tanto de P1 quanto de P2 para a medida de avaliação proposta. As CMs cujo grau de composicionalidade pretendemos aferir foram extraídas das 10 listas apresentadas no capítulo 3 e ranqueadas em ordem crescente de transparência semântica. As Tabelas abaixo (Tabelas de 4 a 13) estão organizadas da seguinte forma:

– cada uma delas contém CMs retiradas de cada uma das 10 listas de CMs apresentadas no capítulo 3: *fazer+SN, ter+SN, dar+SN, perder+SN, usar+SN, receber+SN, deixar+SN , tomar+SN, ganhar+SN, criar+SN.*

– a coluna da extrema esquerda contém a lista de SNs na estrutura V+SN;

– *SM1* é a similaridade média intra-*P*1, ou seja, entre as CMs;

– *SM2* é a similaridade média intra-*P*2; ou seja, entre os SNs;

– *SM3* é a similaridade média entre *P*1 e *P*2*; Var* são as variâncias correspondentes.

Portanto, quanto maior for o valor de *SM3*, mais similares são os parágrafos contendo a CM e os parágrafos contendo o SN fora da CM. Conseqüentemente, mais composicional ou transparente será a CM. Além disso, quanto maior o grau de similaridade entre as CMs (SM1), menos polissêmica é a CM. Quanto maior o grau de similaridade entre os SNs (SM2), menos polissêmico é o SN.

É importante ressaltar, nesta etapa do estudo, que para os casos em que as ocorrências de CMs não eram suficientes para uma aplicação confiável do método (menos de 30 ocorrências), nós adicionamos um córpus através da ferramenta *Google*™ através de buscas contendo as CMs de mesmo padrão tanto na forma canônica quanto na forma flexionada[20].

FAZER	SM1	Var	SM2	Var	**SM3**	Var
falta	1,85	0,01	5,59	0,10	**0,26**	0,0003
a festa	0,77	0,004	1,95	0,004	**0,33**	0,0003
sentido	5,33	0,07	3,8	0,04	**0,33**	0,0005
água	3,13	0,04	7,81	0,02	**0,35**	0,0005
dinheiro	0,64	0,001	2,98	0,007	**0,36**	0,0005
amigos	0,62	0,0008	3,5	0,02	**0,36**	0,0007
compras	1,44	0,013	2,03	0,007	**0,44**	0,002
parte	0,46	0,0004	2,72	0,09	**0,5**	0,003
sucesso	0,84	0,006	2,52	0,04	**0,5**	0,0007
campanha	0,60	0,003	2,45	0,016	**0,54**	0,0004

Tabela 4: resultados com verbo *fazer*

[20] As buscas eram feitas através de escolhas aleatórias das flexões dos verbos.

TER	SM1	Var	SM2	Var	**SM3**	Var
fôlego	1,066	0,01	2,17	0,005	**0,34**	0,0005
acesso	0,86	0,007	3,04	0,01	**0,36**	0,0002
uma idéia	0,54	0,008	1,94	0,009	**0,36**	0,0004
razão	1,00	0,007	5,29	0,12	**0,42**	0,001
sucesso	1,47	0,03	4,11	0,04	**0,43**	0,0007
força	5,98	0,13	4,06	0,006	**0,44**	0,0004
problema	0,88	0,002	2,51	0,02	**0,72**	0,001
medo	1,28	0,017	5,42	0,07	**0,84**	0,006

Tabela 5: resultados com o verbo *ter*

DAR	SM1	Var	SM2	Var	**SM3**	Var
bandeira	1,53	0,0290	1,69	0,002	**0,19**	0,0001
frutos	2,54	0,0570	3,13	0,010	**0,28**	0,0003
tempo	2,16	0,0600	6,15	0,270	**0,28**	0,0004
sorte	0,80	0,0010	4,40	0,060	**0,62**	0,0030
entrevistas	0,86	0,0030	12,10	0,400	**0,66**	0,0006
resultado	0,87	0,0070	4,33	0,120	**0,69**	0,0005
lucro	5,20	0,1500	5,07	0,100	**0,94**	0,0010
declarações	1,01	0,0009	18,67	0,880	**1,07**	0,0020

Tabela 6: resultados com o verbo *dar*

PERDER	SM1	Var	SM2	Var	**SM3**	Var
a cabeça	0,79	0,004	1,24	0,005	**0,18**	0,0002
o bonde	1,12	0,02	1,91	0,003	**0,37**	0,0005
peso	3,08	0,02	5,18	0,01	**0,48**	0,0008
dinheiro	0,57	0,0007	7,72	0,12	**0,57**	0,001
tempo	0,39	0,0004	5,31	0,05	**0,58**	0,002
a eleição	1,39	0,002	5,67	0,07	**0,74**	0,0009
o emprego	1,25	0,01	3,15	0,01	**1,04**	0,02

Tabela 7: resultados com o verbo *perder*

USAR	SM1	Var	SM2	Var	**SM3**	Var
a cabeça	0,91	0,005	2,35	0,005	**0,259**	0,0003
a força	7,63	0,18	2,26	0,005	**0,37**	0,0006
o cinto	19,71	2,09	5,59	0,08	**0,55**	0,0008
computador	1,14	0,002	4,52	0,013	**0,63**	0,0003
drogas	2,35	0,007	3,98	0,015	**0,65**	0,0007
camisinha	2,09	0,03	6,12	0,17	**0,76**	0,001

Tabela 8: resultados com o verbo *usar*

RECEBER	SM1	Var	SM2	Var	**SM3**	Var
alta	2,43	0,008	17,0	0,3	**0,41**	0,0003
visita	1,15	0,0006	10,8	0,032	**0,6**	0,0004
bola	1,94	0,05	8,07	0,16	**0,70**	0,0006
dinheiro	0,69	0,0007	4,93	0,067	**0,75**	0,005
benefício	2,15	0,008	4,04	0,1	**1,13**	0,001
propina	1,91	0,004	12,40	0,24	**1,45**	0,001

Tabela 9: resultados com o verbo *receber*

DEIXAR	SM1	Var	SM2	Var	**SM3**	Var
marcas	1,129	0,005	2,82	0,008	**0,31**	0,0003
o país	0,67	0,0007	2,49	0,02	**0,5**	0,003
o cargo	0,68	0,003	2,32	0,02	**0,51**	0,002
o governo	0,65	0,0005	3,67	0,05	**0,55**	0,0005
o local	0,64	0,05	3,67	0,54	**0,55**	0,0005
a cidade	0,67	0,0004	4,02	0,03	**0,60**	0,0004
os filhos	1,18	0,01	6,04	0,04	**0,65**	0,001
vestígio	8,33	0,33	10,61	0,2	**0,7**	0,0004

Tabela 10: resultados com o verbo *deixar*

TOMAR	SM1	Var	SM2	Var	**SM3**	Var
partido	0,96	0,001	3,18	0,02	**0,3**	0,0003
iniciativa	1,76	0,004	12,9	0,60	**0,31**	0,0002
conhecimento	1,0	0,007	2,13	0,006	**0,33**	0,006
café	1,8	0,02	27,8	3,53	**0,4**	0,0005
o poder	0,80	0,007	2,41	0,003	**0,47**	0,007
posse	1,14	0,009	3,41	0,04	**0,52**	0,0004
uma decisão	0,45	0,0005	1,77	0,012	**0,54**	0,002
banho	1,32	0,01	3,0	0,03	**0,8**	0,01
providência	1,11	0,01	3,15	0,03	**0,94**	0,01
cuidado	0,9	0,01	6,45	0,08	**0,99**	0,01

Tabela 11: resultados com o verbo *tomar*

GANHAR	SM1	Var	SM2	Var	**SM3**	Var
espaço	4,64	0,12	2,51	0,01	**0,25**	0,0002
terreno	1,13	0,006	2,29	0,07	**0,28**	0,0002
força	7,76	0,18	3,48	0,01	**0,31**	0,0003
dinheiro	0,5	0,001	3,28	0,06	**0,31**	0,001
tempo	0,54	0,001	2,56	0,02	**0,48**	0,0009
a vida	0,71	0,002	8,59	0,15	**0,62**	0,001
o jogo	1,05	0,001	6,23	0,08	**0,67**	0,001
a eleição	1,05	0,05	6,3	0,13	**0,69**	0,0006

Tabela 12: resultados com o verbo *ganhar*

CRIAR	SM1	Var	SM2	Var	**SM3**	Var
raízes	2,43	0,008	17,0	0,3	**0,41**	0,0003
atritos	1,15	0,0006	10,8	0,032	**0,6**	0,0004
polêmica	1,94	0,05	8,07	0,16	**0,70**	0,0006
obstáculos	0,69	0,0007	4,93	0,067	**0,75**	0,005
ameaças	2,15	0,008	4,04	0,1	**1,13**	0,001
os filhos	1,91	0,004	12,40	0,24	**1,45**	0,001

Tabela 13: resultados com o verbo *criar*

4.3
Avaliação dos resultados

A Tabela 14 abaixo apresenta os resultados qualitativos dos testes aplicados com cada verbo. Na coluna da extrema esquerda figuram as CMs consideradas como as mais composicionais ou transparentes em relação às outras encabeçadas pelo mesmo verbo. Isto é, aquelas que mais apareceram em microcontextos similares aos microcontextos dos SNs que as compõem. A coluna da extrema direita apresenta as CMs menos composicionais, ou mais semanticamente opacas. Ou seja, aquelas que menos ocorreram em microcontextos similares aos microcontextos dos SNs que as compõem. Já na coluna do meio figuram os casos não-extremos ou os meio-termos.

Lema	***+ composicional***	***meio-termo***	***- composicional***
FAZER	fazer sucesso fazer campanha	fazer compras	fazer falta fazer sentido
TER	ter medo ter problema	ter força	ter fôlego ter uma idéia
DAR	dar lucro dar declarações	dar sorte	dar bandeira dar frutos
PERDER	perder emprego perder a eleição	perder tempo	perder a cabeça perder o bonde
USAR	usar camisinha usar drogas	usar o cinto	usar a cabeça usar a força
RECEBER	receber propina receber benefício	receber visita	receber alta
DEIXAR	deixar vestígio deixar filhos	deixar o país	deixar marcas
TOMAR	tomar cuidado tomar providência	tomar decisão	tomar partido tomar iniciativa
GANHAR	ganhar a eleição ganhar o jogo	ganhar tempo	ganhar espaço ganhar terreno
CRIAR	criar os filhos	criar obstáculos	criar raízes

Tabela 14: resumo qualitativo dos resultados

Os resultados parecem ser, em alguma medida, consistentes com nossas inevitáveis previsões, fornecendo base empírica a nossas parcas intuições sobre os padrões de composicionalidade de CMs do PB. De fato, nós previámos que *tomar partido* e *partido* assim como *dar bandeira* e *bandeira* iriam figurar em microcontextos pouco relacionados entre si; ou seja, microcontextos com um baixo índice de palavras idênticas. Seguem alguns poucos exemplos retirados de *P1* e *P2*, respectivamente, que demonstram esta opacidade semântica entre os microcontextos acima:

Fragmento de P1 *(tomar partido)*
"*Se queremos agir, temos de* ***tomar partido****, «sujar as mãos», e não só no sangue, que é nobre, mas também «na merda» nas alianças sujas, na mentira"* .
" Era uma coisa complicada, que minha mãe não aceitava de vez em quando explodia, a gente via as consequências, tinha que ***tomar partido****, isso ao longo de anos."*
"Em vez de ***tomar partido****, a mente independente de Einstein preferiu encarar os paradoxos entre as duas correntes e tentar unificá-las."*.

Fragmento de P2 *(partido*)
*"O **partido** vai presidir as Comissões de Trabalho, Administração e Serviço Público, a de Seguridade Social e Família e, por fim, a Comissão de Defesa Nacional".*
*"«Quero ver o Vernon Reid dar um» break» no rock e tocar maracatu e **partido** alto aí no Brasil», disse ontem Naná Vasconcelos, por telefone, de Nova York."*
*"Depois de ter se assumido como bom **partido**, solteiro, em busca de namorada e posado até de cuecas na revista «Caras», o deputado Robson Tuma mais conhecido como Tuminha não tem do que reclamar ."*

Quadro 1: microcontextos de *partido / tomar partido*

Fragmento de P1 (*dar bandeira*)
*"Tem gente que anda com vidrinhos de álcool no carro ou, como a mulher está viajando, leva outra camisa e troca para não **dar bandeira**. "*
*"Dar pala: sinônimo de **dar bandeira** ."*
*"Entrou na Faap, fez cinema, publicidade, teatro, **deu bandeira** ."*
*"Progresso -- Não falei, Ordem, pra você não ficar **dando bandeira** ?"*

Fragmento de P2 *(bandeira)*
*"O verde e amarelo do Brasil ganham o vermelho da **bandeira** do Olodum, que apresentará seu filhote Dança Olodum! , um desdobramento do Bando de Teatro do grupo baiano" .*
*"A **bandeira** 2 está autorizada a partir das 6h de hoje.".*
*"Ainda em maio, o Fashion Mall lança um cartão de crédito internacional de afinidade com a **bandeira** Visa."*
*" Irracionalmente, queimaram a **bandeira** do patrocinador, como se este tipo de pressão trouxesse resultados*".

Quadro 2: microcontextos de *bandeira / dar bandeira*

Por outro lado, também suspeitávamos que *tomar banho* e *banho* assim como *usar camisinha* e *camisinha* ocorreriam em parágrafos similares; isto é, que compartilhassem um grau elevado de palavras iguais. Sobre o primeiro par, estávamos parcialmente certas, uma vez que a CM também é utilizada como uma forma de insulto:

Fragmento de P1 *(tomar banho)*
"Segundo os pesquisadores, os pacientes podem ter sido infectados com microorganismos ao ***tomar banho*** *ou beber água ."*
"Quanto à expressão «ele que vá ***tomar banho****», embora não registrada em fita, foi pronunciada diante de testemunhas que também a interpretaram como sendo dirigida ao ministro Ricupero, e não à reportagem da Folha."*
"Elas podem dormir, ***tomar banho*** *e se alimentar entre 20h30 e 8h. Mas não temos estrutura para atender durante o dia», diz Maria Cecília ."*
"Dentro dos grupos eram combinados esquemas de revezamento, para todos poderem ir para casa almoçar e ***tomar banho****, por exemplo."*

Fragmento de P2 *(banho)*
"O corpo de Betty, na cena do ***banho,*** *está um luxo, e a produção conseguiu um charme a mais: uma paisagem urbana que pode ser de qualquer cidade do planeta".*
" Sua rotina inclui nutrição, ***banho,*** *medicação e fisioterapia."*
" Enquanto Hargreaves e sua irmã Ruth, assessora especial do presidente, vestiam roupas de ***banho,*** *Itamar trajava calça escura e camisa de manga comprida ."*

Quadro 3: alguns microcontextos de *banho/ tomar banho*

Já o segundo par parece estar intimamente relacionado; ou seja, a CM *usar camisinha* aparece em microcontextos que compartilham um grau elevado de palavras iguais. Portanto, dentro da nossa proposta, essa CM parece ser bastante composicional. Além disso, o SN *camisinha* parece apresentar um grau baixíssimo de polissemia, uma vez que SM2, ou seja, a similaridade média entre todos os parágrafos que contêm o SN é altíssimo (6,12).

Fragmento de P1 *(usar camisinha)*
"É preciso dizer ao adolescente que tem de ***usar camisinha****"*
"Se estão com Aids, eles informam o freguês e ou parceiro e pedem para ***usar camisinha..***"
"Você não ia ser louca de ter vida sexual sem ***usar camisinha****, não é verdade ?"*

Fragmento de P2 *(camisinha)*
"Bispos aceitam ***camisinha*** *no combate a Aids* ***."***
"Sem ***camisinha*** *não dá, afirma o analista de sistemas Jorge Luis, 24, que costuma sair com esses adolescentes* ***."***
"*A* ***camisinha*** *foi apontada como o melhor método anticoncepcional ."*

Quadro 4: alguns microcontextos de *usar / usar camisinha*

Por outro lado, suspeitávamos que o par "tomar café" e "café" seguiria o mesmo padrão composicional de "usar camisinha", o que não ocorreu, uma vez que o SN "café" em P2 ocorrem em ambientes lingüísticos não relacionados a P1, tais como economia, agricultura, arquitetura (como uma cor).

Fragmento de P1 *(tomar café)*
*"Ele parou em dois bares, para **tomar café** e água gelada."*
*"Não há mais o risco de sair para **tomar café** e descobrir, de última hora, que a máquina está quebrada ".*
*"Os dois deverão **tomar café** juntos a partir das 8h no Othon Palace Hotel, em frente à praia de Copacabana (zona sul)" .*

Fragmento de P2 *(café)*
*"Chegaram a brincar com os rivais - em sua maioria tensos, apesar do encontro ter acontecido na porta de um **café**, o Cabalas" .*
*Junte-se a todo esse clima o cardápio que inclui frutas tropicais, **café**, caldo de feijão e caipirinha .*
*«Nosso objetivo é impedir a saída clandestina do **café** que nós produzimos e desmantelar as quadrilhas de sonegadores .*
*"De manhã, depois da toilette e do **café**, sentava-se no divã da sala principal e lia os jornais ".*
*"Um conjunto de **café** de US$ 3.000 não é só para coleção ?"*
*"(...) Soldados marcham decididos no rumo da estação carregando acima dos bonés, em meio ao espinhal de canos com as bocas voltadas para o sol, (…) um pedaço de fumo de rolo no bolso inferior e coroando tudo o imenso chapéu cor de **café**, (…).*

Quadro 5: alguns microcontextos de *café/ tomar café*

Em suma, esses resultados nos deixaram extremamente confiantes com a aplicabilidade do método. Portanto, ao invés de atribuir à intuição do pesquisador o poder de aferição do grau de composicionalidade de uma CM, nós preferimos confiar naquilo que o córpus nos revela.

De fato, reconhecemos que os resultados aqui obtidos deveriam ser corroborados por um córpus ainda mais robusto do PB, pois alguém poderia refutar nossas conclusões caracterizando o córpus como tendencioso. Por ora, nós preferimos pensar, ao contrário, que nossa intuição é que tende a ser traiçoeira.