

### 3

## Por um caminho não-representacionista para a detecção dos multivocábulos

*We begin to feel, or ought to, terrified that maybe language (and understanding and knowledge) rests upon very shaky foundation — a thin net over an abyss*

S. Cavell

Neste terceiro capítulo da tese nos aventuramos por uma vereda relativamente pouco trilhada nos estudos lingüísticos, mas que já havia sido sinalizada pelas considerações sofistas da Grécia antiga. Trata-se de um olhar sobre o fenômeno da linguagem que privilegia o seu estatuto convencional. Dispensa, em contrapartida, qualquer tipo de teor simbólico ou representativo da linguagem em relação à realidade ou à mente.

Seguiremos aqui a mesma organização do capítulo anterior. Reservamos a primeira seção para breves considerações filosóficas sobre essa linha de pensamento; apresentamos na segunda seção o seu encaminhamento no âmbito lingüístico e no âmbito de PLN; já na terceira seção, lançamos esse olhar sobre o fenômeno da CM, apresentando a vantagem ou, talvez seja melhor dizer, a conveniência desse ponto de vista se comparado às perspectivas representacionistas discutidas no capítulo 2. Propomos, finalmente, em alinhamento com essa visão, uma avaliação de base estatística para detecção de CMs do tipo V+SN.

### 3.1

#### A herança filosófica

*O discurso é um grande soberano, que com o mais diminuto e inaparente corpo as mais divinas obras executa.*

Górgias, *Elogio a Helena*

Não é tarefa simples caracterizar as considerações não-representacionistas sobre a linguagem. Na história do pensamento filosófico, a visão sofista sobre o

discurso é talvez a primeira manifestação documentada dessa concepção de linguagem. Muito do que sabemos sobre o pensamento sofista em relação à linguagem deve-se aos escritos platônicos.<sup>10</sup> Platão apresenta as considerações sofistas — em que a linguagem é vista à luz da célebre máxima de Protágoras, “o homem é a medida de todas as coisas” — como contraponto ao pensamento socrático. Este formato dialógico apresenta, de certa forma, um teor didático para reafirmação de sua crença representacionista. Em outras palavras, o fato de Platão sustentar, em oposição violenta ao ideário sofista, a existência de verdades únicas e fixas, em alguma medida pode ter minimizado ou desqualificado a importância e teor das considerações sofistas ao longo da história do pensamento filosófico (cf. Marcondes, 1997 e Martins, 2005). Logo, esse investimento de Platão contra os sofistas deve ter contribuído para minimizar e desqualificar também o ângulo anti-representacionista sobre a linguagem.

Dentre as características privilegiadas por essa visão de linguagem está a mutabilidade do significado de uma expressão em virtude de esta depender das práticas humanas, como também a incapacidade de a expressão representar algo exterior à linguagem.

Esse viés pragmático sobre a linguagem é retomado já na história mais recente da filosofia<sup>11</sup>, notadamente, por Ludwig Wittgenstein<sup>12</sup>. De forma ostensivamente resumida, pode-se dizer que o filósofo austríaco se recusa a enxergar o estudo da linguagem dentro dos mesmos moldes metafísicos propostos para as ciências, principalmente porque o “cientista”, neste caso, é uma das peças envolvidas no que chama de “jogos de linguagem” (Wittgenstein, 1979; Glock, 1996). Ele desenvolve este conceito através de uma analogia com a idéia de “jogo”, e aponta certas características e semelhanças com a linguagem: jogos possuem regras, são práticas compartilhadas por uma comunidade, possuem peças, são autônomos, não requerem justificativas. Sua proposta é a de que os jogos de linguagem “são a totalidade da linguagem e das atividades com as quais está interligada” (Wittgenstein, 1979: §§ 7 e 23). Os jogos de linguagem seriam atividades autônomas que prescindem de

<sup>10</sup> O *Sofista* e *Crátilo* são exemplos de diálogos em que Platão tematiza a linguagem.

<sup>11</sup> Para uma boa apreciação sobre o assunto ver Martins, 1999.

<sup>12</sup> Refiro-me ao segundo Wittgenstein e sua visão pragmática expressa em *Investigações Filosóficas*, São Paulo Cultural, Coleção **Os Pensadores**, 1979.

explicação; “são parte de nossa história natural, assim como andar, comer, jogar, etc”. (ibid.: §25).

E o fato de o próprio homem estar encarcerado nos rituais lingüísticos que regem toda e qualquer manifestação lingüística compromete o seu julgamento teórico ou sua tentativa de explicação acerca do assunto. O estudo sobre a linguagem, portanto, seria ele mesmo mais um desses jogos. E como o significado de uma palavra encontra-se na execução da linguagem, não há nada a ser provado ou justificado. Não se pode explicar o que está explícito. A significação é habilidade em lidar com as palavras na linguagem. A explicação é resultado da nossa “ânsia de generalidade”. Ela pode até ser utilizada, mas não deveria ser entendida como uma meta-regra; ou uma regra fora do jogo. Toda explicação é interessada. (ver Scherer, 2002 sobre o conceito de *explicação*).

Em *Da certeza* (§559 in Sumares, 1994) — obra em que o autor tematiza a questão das regras que se cristalizam e passam a funcionar como pressupostos para a construção do conhecimento e como base para nossas ações — ele retoma o conceito de jogo: “O jogo de linguagem é, por assim dizer, imprevisível. Quero dizer: não está fundamentado. Não é racional (ou irracional). Está aí - como a nossa vida”.

Sua visão nega, deste modo, a vocação representacionista da linguagem tão defendida por filósofos como Platão, Aristóteles e Locke. Não há, para ele, uma essência do significado que preceda o uso das palavras. É somente na prática do uso que o significado se dá. E mesmo assim, ele não poderia ser “coisificado”. Por essa razão, é instável, contingente e fragmentado. Falar sobre a linguagem, ou fazer um relato sobre o relato, será um relato do jogo, e também faz parte dele. Portanto, nunca será definitivo.

Se num primeiro momento sua visão parece amputar a Lingüística, e mais especificamente, eliminar a Semântica, uma vez que sugere uma total impossibilidade de tratamento do fenômeno da significação, concordamos com Martins (1999:147) quando pondera que adotar uma visão wittgensteiniana na Lingüística “não corresponde à negação da possibilidade de qualquer estudo sistemático e empírico das línguas do mundo”. A autora sugere, portanto, que uma Lingüística sob a perspectiva wittgensteiniana “é viável e teria como propósito geral descrever as regularidades — parciais e contingentes —

observáveis nos jogos de linguagem que constituem as línguas do mundo”(idem).

A descrição de uma língua, segundo o filósofo, estará sempre condicionada à parcialidade imposta pelas infindáveis possibilidades de lances dos jogos de linguagem. Sobre a completude da linguagem, Wittgenstein confronta o leitor com o seguinte questionamento no §18 de *Investigações Filosóficas*:

][...] (e com quantas casas ou ruas, uma cidade começa a ser cidade?) Nossa linguagem pode ser considerada como uma velha cidade: uma rede de ruelas e praças, casas novas e velhas, e casas construídas em diferentes épocas; e isto tudo cercado por uma quantidade de novos subúrbios com ruas retas e regulares e com casas uniformes.

Esse parágrafo de *Investigações Filosóficas* é particularmente relevante para o presente estudo porque se alinha à nossa escolha teórica e também metodológica: uma abordagem com base em córpus. Um exemplário da língua sempre pode ser maior, é verdade; mas nunca deixará de ser um extrato válido do seu uso. Portanto, a idéia de abarcar a língua de uma forma totalizante é uma utopia. Conclui-se, daí, que a parcialidade é inevitável.

Ao longo do tempo, algumas expressões são construídas, outras demolidas. Isto é parte do jogo. Todo córpus de uma língua seria, portanto, uma caracterização válida da mesma<sup>13</sup>. Voltaremos a esta questão mais adiante na seção 3.3.

### 3.2

#### **Ecos do não-representacionismo na Lingüística e em PLN**

Na falta de uma explicação satisfatória para a noção do significado, os lingüistas que atuam na área da Semântica encontram-se na situação de não saber do que é que estão falando

W. Quine

A posição do lingüista Roy Harris (1996), em relação ao funcionamento da língua alinha-se às idéias não-representacionistas solidificadas por

<sup>13</sup> para uma melhor apreciação das idéias wittgensteinianas ver Glock, 1996; Martins, 1999.

Wittgenstein. Em Scherer (2002) é possível encontrar uma farta exposição da simpatia de Harris às idéias do filósofo austríaco. Um dos pontos em comum entre as duas posições é a impossibilidade de distinção entre semântica e pragmática, ou entre conhecimento lingüístico e enciclopédico.

Note-se que a visão pragmática oferecida aqui se distancia de muitas abordagens pragmáticas da Lingüística. Estas conduzem suas explicações sobre o “uso da língua” com base na distinção *Semântica-Pragmática*, que com freqüência se apóia na distinção *significado literal-significado de uso*; algo indissociável por uma visão não-representacionista. Scherer (2002:26) avalia, portanto, que Wittgenstein e Harris compartilham uma visão pragmática radical, em que o uso lingüístico não é um dos componentes da linguagem, mas a única forma produtiva de se pensar os fenômenos lingüísticos.

Adam Kilgarriff, estudioso do léxico do ponto de vista computacional, também advoga este pragmatismo radical. Em seu contundente artigo “*I don’t believe in word senses*” (2000), ele atribui ao córpus o poder de desambiguação de significados. Demonstra que as palavras, em suas diversas acepções, estão desatreladas do seu alegado “sentido atômico” (*core meaning*). Propõe, como alternativa, uma ontologia, ou um conjunto de palavras semanticamente relacionadas, com base em uma convergência (*cluster*) estatística dos seus vários usos em córpus, em detrimento do seu sentido atômico. Os extratos do córpus são agrupados em sentidos específicos de acordo com os objetivos de tarefas específicas. Acredita, portanto, que os significados só existem dependentes de propostas ou tarefas (*tasks*) Ele conclui: “*in the absence of such purposes, word senses do not exist*”.

A visão de CMs proposta aqui insere-se neste pragmatismo radical. Dessa forma, o intuito é caracterizar as CMs com base no seu uso. É bem verdade que ao optarmos por considerar somente as CMs do tipo V+SN, estaríamos inevitavelmente reificando ou coisificando esse padrão de fragmento lingüístico. Mas é essa a nossa “tarefa”, pelo apelo prático desse padrão estrutural tão recorrente no PB.

A abordagem que seguimos, portanto, é, em parte, filiada àquela proposta por Manning & Schütze (2003), pois, assim como esses estudiosos de PLN, estamos priorizando a freqüência das CMs e iremos utilizar o mesmo método

estatístico proposto por eles para a sua detecção. Os autores também creditam sua proposta a uma inspiração wittgensteiniana (p.17):

*Philosophically, this brings us close to the position adopted in Wittgenstein (that is, Wittgenstein, 1968), where the meaning of a word is defined by the circumstances of its use (a use theory of meaning)[...]Under this conception, much of NLP research directly tackles questions of meaning.*

Diferentemente de Manning & Schütze, no entanto, não faremos nenhum tipo de caracterização apriorística do grau de opacidade semântica das CMs, que os autores rotulam por *colocações*: “uma combinação freqüente na língua com sentido específico entre, no mínimo, duas palavras — um bigrama — abrangendo expressões *totalmente opacas*, cujos itens não contribuem para o significado total da expressão (expressões idiomáticas), ou cujos itens *contribuem parcialmente* para o significado total da expressão” (p. 141, grifo nosso). Acreditamos que essa definição implica uma abordagem representacionista da linguagem. Propomos, em contrapartida, um novo tipo de aferição do grau de composicionalidade, guiada por medidas de similaridade entre adjacências lingüísticas, explicitada no capítulo 4.

### 3.3

#### A inevitabilidade do paradoxo do córpus

Em um estudo lexicográfico para o francês, Verlinde & Selva (2001) compararam a abordagem tradicional da construção de um dicionário de aprendizes de segunda língua, baseada na intuição do lexicógrafo, a uma abordagem baseada em córpus. De um modo geral, verificaram que é atribuído ao número de entradas (macroestrutura) de um dicionário uma importância maior do que ao conteúdo de cada entrada (microestrutura). Portanto, a ênfase recai nas palavras simples em detrimento das combinações de palavras, o que vem a ser um contra-senso, já que, para fins comunicativos, os aprendizes necessitam muito mais de informações sobre combinações do que sobre palavras isoladas.

Os pesquisadores puderam detectar, por exemplo, que, mesmo com a difundida repressão do governo francês aos anglicismos, o córpus apresentava

uma frequência bastante representativa desses estrangeirismos. Os autores, no entanto, utilizaram como fontes periódicos jornalísticos, o que, segundo John Sinclair, pode ser questionável: “um *córpus* é uma coletânea de um material amplamente homogêneo, mas retirado de fontes diversas de forma que a individualidade de uma fonte se perde, a não ser que o pesquisador queira isolar um texto em particular. A diversidade de fontes é uma garantia de segurança dos dados” (Sinclair, 1991: 17-18).

Dentro dessa mesma perspectiva teórica, Verlinde & Selva (2001) destacam que a lexicografia de *córpus* é uma evidência empírica necessária à intuição do pesquisador, que serviria para preencher as lacunas do *córpus* que chamam de “não equilibrado”. Portanto, eles ressaltam a urgência da construção de um *córpus* equilibrado para o francês e outras línguas.

Podemos notar em grande parte dos lexicógrafos que se dedicam ao estudo de *córpus* o sentimento de que a compilação de fontes diversas evita o paradoxo do *córpus*. Contudo, essa caracterização do *córpus* ideal ignora a imagem da “cidade” apresentada por Wittgenstein e revela uma ilusão corrente na Lingüística: a de que é possível fugir do paradoxo do *córpus*. A questão intocada é a de que a língua em si não é completa; sempre é possível acrescentar mais uma casa ou mesmo uma rua, o que torna o *córpus* um fragmento de algo já fragmentado. Portanto, não há como fugir do paradoxo do *córpus* uma vez que a completude da língua também é algo inatingível.

Tal constatação, entretanto, não desabona um estudo descritivo da língua; o fato de não ser possível exaurir todas as possibilidades de jogos da linguagem não impede um lingüista ou um lexicógrafo de descrever o que geralmente é constatado nos jogos (cf. Martins, 1999: 147). Por outro lado, é desejável que o pesquisador se cerque de alguns critérios para aferir a adequação de um *córpus* em função dos objetivos pragmáticos que se deseja alcançar.

Por essa razão, dentre um escopo restrito de opções de *córpus* significativos no PB, recorreremos a um *córpus* jornalístico. Concluimos que seria um extrato de língua atraente para a nossa tarefa.

### 3.3.1

#### O **cópus utilizado: CETENFolha**

O **CETENFolha** (Corpus de Extractos de Textos Electrónicos NILC/Folha de S. Paulo) é um **cópus** jornalístico de cerca de 24 milhões de palavras em PB, parte integrante do **cópus** NILC<sup>14</sup>, fornecido pelo projeto Linguateca ([www.linguateca.pt](http://www.linguateca.pt)). A *Folha de S. Paulo* é um jornal diário brasileiro de grande circulação. Além das habituais seções diárias, o jornal contém igualmente diversos cadernos não-diários, a maior parte dos quais foram incluídos no **cópus**. Existe também desde 1996 uma edição eletrônica (<http://www.folha.uol.com.br>). Dentre as seções incluídas no **cópus** CETENFolha estão, por ordem alfabética, Agrofolha, Brasil, Caderno Especial, Cotidiano, Dinheiro, Empregos, Esporte, Folha Ciência, Folhateen, Folhinha, Fovest, Ilustrada, Imóveis, Informática, Mais!, Mundo, Opinião, Revista Folha, TV Folha, Tudo, Turismo, Veículos.

Não foram incluídos no **cópus**:

- os artigos de primeira página que apenas chamam o artigo principal nas páginas interiores: *NORDESTE Chuvas voltam ao sertão da Paraíba e trazem esperança aos agricultores PÁG. 3*;
- os artigos com menos de 250 caracteres;
- algumas chamadas para outras páginas dentro de um artigo: (*leia mais na página 3*);
- anotações internas que se considerou não fazer parte de notícias ou outro texto do jornal, e que não tinham relevância para o **corpus**: (*Edição: São Paulo*);
- artigos duplicados (i.e., quando existiam duas cópias do mesmo artigo, só uma foi incluída).

A nosso ver, portanto, a predominância do discurso semi-formal, a transcrição de discursos diretos em entrevistas e a variedade de assuntos disponibilizada pelas diversas seções do jornal tornam esse gênero uma fonte significativa de jogos de linguagem no PB.

---

<sup>14</sup> O **corpus** NILC -USP, contém textos brasileiros do registro jornalístico, didático, epistolar e redações de alunos. Trata-se de um extrato com 37 milhões de palavras.

De fato, a opção por um *cópus* de teor jornalístico tem suas implicações: a língua fica prioritariamente associada àquilo que é considerado notícia em detrimento, por exemplo, de um bate-papo despretensioso entre adolescentes. Entretanto, a escolha por esse tipo de extrato da língua também está associada à falta de um *cópus* mais robusto do PB. Uma outra razão da opção pelo *cópus* CETENFolha está no fato de ser o único significativo no PB disponível para *download*; e, portanto, o único passível de aplicação dos testes probabilísticos que virão mais adiante.

### 3.4

#### O teor estatístico do fenômeno lingüístico

Embora proposta antes do advento de *cópus* computadorizados, a Teoria Contextual do Significado proposta pelo lingüista britânico J. R. Firth (1957), subjaz a uma abordagem a partir de *cópus*. Ele propõe que o estudo do significado e do contexto devem ser centrais para a lingüística. Recusa-se a admitir qualquer tipo de distinção entre *langue* e *parole* (estabelecida antes dele por Saussure) ou entre *competência* e *desempenho* (estabelecida mais tarde por Chomsky), já que não enxerga a língua como uma entidade autônoma. Os *eventos lingüísticos*, nos seus termos, são recorrentes e repetidamente observáveis. Dessa forma, sua visão se alinha à perspectiva pragmática adotada neste estudo — que, por sua vez, leva a uma perspectiva computacional, já que só podemos constatar mais precisamente certos eventos na língua com auxílio de um aparato que possibilite o armazenamento desses dados e sua recuperação.

No mesmo ano em que Firth profetizou “*You shall know a word by the company it keeps*”, Chomsky tornou pública, em *Syntactic Structures*, a sua crença de que a língua deveria ser analisada em um nível muito mais profundo do que sua estatística superficial. Isto é, numa crítica à visão empirista, argumentou que uma abordagem com base em dados da língua consideraria como identicamente remotas as sentenças “*Colorless green ideas sleep furiously*” e “*Furiously sleep ideas green colorless*”. Ou seja, em qualquer modelo estatístico, as duas sentenças seriam igualmente excluídas, embora, como falantes, nós saibamos que a primeira obedeceria a algumas regras

gramaticais. O que Chomsky defendia é que a primeira não poderia ser totalmente desconsiderada e equiparada à segunda simplesmente pelo fato de sua probabilidade de ocorrência ser remota<sup>15</sup>. Essa crítica, conhecida como *problema dos dados esparsos*, foi altamente influente para uma mudança de perspectiva teórica nos anos 60.

Mas a evolução de técnicas estatísticas aliada à possibilidade de construção de *cópus* mais robustos de alguma forma equacionou aquilo que Chomsky utilizava como emblema de sua crítica. Além disso, a forma de Firth enxergar a língua não se presta à comparação com o modelo chomskyano. Para ele a língua é um *evento*, uma forma de ação (*a way of “doing things”*); e é por isso que seria um ato legítimo por parte do lingüista se ater aos eventos discursivos propriamente ditos. Firth acreditava que o evento era um fim em si mesmo e não uma forma de acessar o “verdadeiro” objeto de estudo: o sistema subjacente a ele, como Chomsky advogava. E o nosso estudo alinha-se a esse perspectiva eventiva da língua.

### 3.4.1

#### Mãos à obra

A seguir está a primeira etapa prática desta pesquisa. Nesta seção apresentamos uma listagem dos verbos mais recorrentes no *cópus* CETENFolha em três situações distintas: sua frequência absoluta (Tabela 1); sua frequência seguida facultativamente de determinante e obrigatoriamente de um nome, formando a estrutura V+(det)+N (Tabela 2); sua frequência seguida facultativamente de determinante e obrigatoriamente de um nome seguido obrigatoriamente de marcas de pontuação ou advérbio ou conectivo, formando a estrutura V+(det)+N intransitiva (Tabela 3). Já em 3.5.1 apresentamos um teste estatístico (Banerjee & Pedersen, 2003), para detectar as CMs, encabeçadas por cada um dos 10 verbos mais frequentes com esse padrão, em

---

<sup>15</sup> Ironicamente, a probabilidade de ambas as frases ocorrerem é grande em virtude de sua relevância ilustrativa na Lingüística.

detrimento de combinações sintáticas aleatórias. Após essa detecção, o método lista as CMs por ordem de frequência.

O critério estatístico para detecção de CMs na língua tem uma dupla função. Primeiramente, pela própria natureza freqüencial do fenômeno, consideramos ser um recurso primordial para trazer à tona os padrões de co-ocorrências mais utilizados; em segundo lugar, ao lançar mão de recursos estatísticos para detecção das CMs mais freqüentes, não apenas poupamos tempo e trabalho de que necessitaríamos para verificar se uma seqüência é ou não freqüente na língua, como também contamos com um valioso aliado para corroborar a listagem: dados reais da língua.

Numa avaliação do cópua CETENFolha, listamos os 30 verbos mais recorrentes. Isto é, sua frequência absoluta:

<b>CETENFolha (24 milhões de palavras)</b>			
<i>ranqueamento</i>	<i>ocorrências</i>	<i>lema</i>	<i>percentual no cópua</i>
1	109282	ser	0.46%
2	39422	ter	0.16%
3	36743	estar	0.153%
4	20668	poder	0.086%
5	20440	ir	0.085%
6	14660	fazer	0.061%
7	12880	haver	0.054%
8	11928	dever	0.049%
9	8941	querer	0.037%
10	7037	dar	0.029%
11	6852	ficar	0.028%
12	6822	vir	0.028%
13	5576	dizer	0.023%
14	5049	chegar	0.021%

15	4836	passar	0.020%
16	4369	saber	0.018%
17	3952	começar	0.016%
18	3818	acontecer	0.016%
19	3523	conseguir	0.015%
20	3483	receber	0.014%
21	3369	ver	0.014%
22	3279	levar	0.013%
23	3041	deixar	0.012%
24	2999	existir	0.012%
25	2901	precisar	0.012%
26	2744	ocorrer	0.011%
27	2734	trabalhar	0.011%
28	2710	sair	0.011%
29	2663	pretender	0.011%
30	2662	ganhar	0.011%

Tabela 1: frequência absoluta dos 30 verbos mais recorrentes no cópuz

De acordo com Zipf (1949) "quanto maior a frequência de uma palavra ou morfema, maior será o número de combinações possíveis (grosso modo, compostos e formas morfológicamente complexas)." Seguindo essa hipótese, portanto, as CMs verbais mais frequentes na língua seriam encabeçadas pelos verbos apresentados na Tabela 1, que são os mais frequentes no cópuz.

Vejamos agora a frequência desses mesmos verbos seguidos facultativamente de determinante e obrigatoriamente de nome, formando a estrutura V+(det)+N<sup>16</sup>. Em negrito estão os verbos que não figuram na primeira tabela.

<sup>16</sup> Tomando como exemplo o verbo *fazer*, o formalismo para tal detecção em *linguateca.com.pt* seria ([lema="fazer" & pos="V"] [pos="DET.\*"]? [pos="N"] [classe="JOCF"]). A fórmula JOCF se refere ao cópuz CETENFolha, em detrimento de outros também disponíveis no Linguatca.

<b>CETENFolha</b> <b>V+(det)+N</b>		
<i>ranqueamento</i>	<i>ocorrências</i>	<i>lema</i>
1	29522	ser
2	24570	ter
3	20101	fazer
4	8545	haver
5	7037	dar
<b>6</b>	<b>4241</b>	<b>usar</b>
7	3483	receber
8	3369	ver
<b>9</b>	<b>3320</b>	<b>criar</b>
<b>10</b>	<b>3237</b>	<b>tomar</b>
11	3035	deixar
<b>12</b>	<b>2888</b>	<b>perder</b>
<b>13</b>	<b>2841</b>	<b>pedir</b>
14	2824	levar
15	2682	pagar
16	2625	ganhar
<b>17</b>	<b>2268</b>	<b>comprar</b>
18	2047	passar
<b>19</b>	<b>1850</b>	<b>vender</b>
20	1528	conseguir
<b>21</b>	<b>1471</b>	<b>tirar</b>
<b>22</b>	<b>1323</b>	<b>causar</b>
23	1286	existir
24	1209	dizer
<b>25</b>	<b>1121</b>	<b>jogar</b>
<b>26</b>	<b>1064</b>	<b>matar</b>
<b>27</b>	<b>1019</b>	<b>virar</b>
28	914	estar
29	866	querer
30	818	começar

Tabela 2: frequência de verbos mais recorrentes do corpus CETENFolha seguidos facultativamente de determinante e obrigatoriamente de um nome.

Note-se que exceto por “ser” e “estar”, a posição dos verbos é bastante diferente da sua frequência absoluta (ver Tabela 1). Além disso, a Tabela 2 apresenta verbos que nem mesmo figuram entre os trinta mais frequentes da Tabela 1 (expostos em negrito).

A Tabela 2, entretanto, não representa a frequência ideal de V+(det)+N com teor intransitivo no *cópus*. Como essa lista se refere a qualquer segmento V+(det)+N, sem qualquer outra restrição de recorrência vocabular, pode incluir também CMs de padrão V+SN+SPrep (*dar o braço a torcer; dar o ar da graça; tirar o cavaleiro da chuva*) assim como V+det+N+pron (*tem gente que adora...*).

A saída menos comprometida para restringir o escopo da busca por uma estrutura V+(det) +N com teor intransitivo foi a inclusão de uma restrição com base em marcas de pontuação, conectivo e advérbio (para que fosse possível, por exemplo, a detecção da expressão *Ele teve alta*; mas também *ele teve alta ontem*; assim como *ele teve alta e já está em casa*)<sup>17</sup>. Abaixo segue a frequência dos verbos mais recorrentes com esse padrão no *cópus*. Em negrito estão os 10 verbos selecionados para implementação dos testes estatísticos com padrão V+(det)+N. Em cinza estão os verbos que foram descartados da implementação estatística, como justificado mais adiante.

CETENFolha		
V+(det)+N com padrão intransitivo		
<i>ranqueamento</i>	<i>ocorrências</i>	<i>lema</i>
1	7511	ser
2	5132	fazer
3	4958	ter
4	1993	dar
5	1749	haver
6	1075	perder
7	1038	usar
8	981	receber
9	940	deixar

<sup>17</sup> Tomando como exemplo, o verbo *ter*, o formalismo de busca seria: ([lema="ter" & pos="V"] [pos="DET.\*"]? [pos="N"] [word="\.,|;|:|\?|!"] | "pos=KC.\*" | "pos=ADV.\*" [classe="JOCF"])

10	<b>901</b>	<b>tomar</b>
11	<b>861</b>	<b>ganhar</b>
12	<b>817</b>	<b>ver</b>
13	<b>770</b>	<b>criar</b>
14	732	pagar
15	728	comprar
16	564	pedir
17	541	levar
18	525	vender
19	480	dizer
20	438	matar
21	423	passar
22	402	causar
23	400	conseguir
24	387	virar
25	375	jogar
26	371	estar
27	342	tirar
28	262	querer
29	249	começar
30	237	comentar

Tabela 3: frequência de verbos seguidos facultativamente de determinante e obrigatoriamente de um nome posposto por marcas de pontuação, conectivo ou advérbio.

Os 30 verbos mais frequentes expostos na Tabela 3 são essencialmente os mesmos da Tabela 2, muito embora o ranqueamento da Tabela 3 não espelhe as ocorrências dos verbos da Tabela 2, exceto pelo verbo “ser”. O único verbo que não figura nessa terceira tabela é “existir”, cujas ocorrências (234) foram excedidas pelo verbo “comentar” (237).

Segundo Vale (2001), a expressão cristalizada mais recorrente em PB é a V+(det)+N de teor intransitivo — que dentro de outra proposta teórica, rotula como V<sub>0</sub>N<sub>0</sub>C. Ele afirma também que os verbos mais recorrentes em CMs em geral, que ele rotula como *Expressões Cristalizadas*, são os verbos-suportes ou verbos leves como “dar”, “ter”, “fazer”, “ficar”, “levar”, “tomar”, “tirar”, “pôr”.

De fato, “dar”, “fazer”, “ter” são verbos que figuram no topo da lista da

Tabela 3. Por outro lado, os nossos resultados desmonstram que há verbos que não se caracterizariam propriamente como leve, mas que são altamente freqüentes com a estrutura V+(det)+N. Por exemplo, “perder” (6°) “usar” (7°) “deixar” (9°), “ganhar” (11°), “criar” (13°), dentre outros, superam a freqüência de verbos tradicionalmente rotulados por leves ou suporte, como “levar”, “tirar”, por exemplo, que estão em 17° e 27°, respectivamente.

Além desses, figuram o verbo de cópula “ser” (1°) e o verbo impessoal “haver” (5°). Como o verbo “ser” além de ser sempre o mais freqüente dos verbos em qualquer cópula de prosa, também apresenta uma farta variedade de combinações estruturais (funcionando quase como uma palavra funcional) optamos por excluí-lo da nossa pesquisa, já que seu ecletismo sintático ou sua explosão combinatória acabaria falseando o teste de verificação de CM. Já o verbo “haver” foi excluído da análise pelo seu teor impessoal. Ao contrário dos outros verbos presentes na Tabela 3, ele prescinde de sujeito (*Houve tumulto; há jurisprudência*), o que o distancia de estruturas de teor agentivo (como, por exemplo N *fazer* N, N *criar* N) ou passivo (como, por exemplo N *receber* N, N *ganhar* N).

Uma outra exclusão também foi feita em relação às CMs encabeçadas pelo verbo “ver” em função de seu teor metalingüístico no cópula, como em *ver tabela, ver resultados, ver fotos, ver textos, etc.* Muito embora a descrição do cópula afirmasse que esse tipo de informação tivesse sido excluída — “algumas chamadas para outras páginas dentro de um artigo” —, isso parece não ter sido feito de forma exaustiva.

Portanto, as 10 estruturas V+(det)+N que serão listadas mais adiante dizem respeito aos verbos que figuram em negrito na tabela 3. São eles: “fazer”, “ter”, “dar”, “perder”, “usar”, “receber”, “deixar”, “tomar”, “ganhar”, “criar”. Note-se que este é um recorte meramente metodológico e que sua arbitrariedade, necessária para os limites desta pesquisa, deixará de avaliar outras estruturas com o mesmo padrão que também são bastante freqüentes com outros verbos.

### 3.5

#### A identificação das CMs

Existem alguns métodos estatísticos disponíveis para a detecção de CMs. Dentre eles, o mais simples e previsível é a seleção baseada na frequência. Consiste na computação das frequências de pares de palavras (bigramas). Os bigramas mais frequentes seriam candidatos naturais a CMs. O grande problema desse método é que as palavras mais frequentes da língua tendem a se combinar mais do que aquelas não tão frequentes (é o caso do verbo “ser”).

No caso das combinações V+det+N, por exemplo, o padrão contém uma palavra funcional: os determinantes. Trata-se de uma classe de palavras com frequência tão elevada que sua combinação com nomes frequentes poderia ser sempre analisada como uma CM. Para que isso não ocorra, aplica-se um filtro de classes de palavras, onde somente se consideram padrões candidatos potenciais a sintagmas. Tal método apresenta algum grau de eficiência; porém, dada a sua simplicidade, os resultados não são considerados suficientemente precisos e destinados somente a combinações fixas (ex: nomes compostos).

No caso de ocorrências descontínuas, ou seja, de combinações mais flexíveis, como é o caso de CMs de base verbal (*fazer muito tempo; tirar todas as dúvidas*), são necessários recursos mais sofisticados, dentre os quais estão métodos como Média & Variância e Testagem de Hipóteses. Este último método é subdividido em Teste t, Testagem de hipóteses de diferenças, Teste  $\chi^2$  de Pearson e Logaritmo de Verossimilhança (Manning & Schütze, 1999: cap. 6)

O método de Média & Variância, contudo, não é destinado para detecção de uma estrutura sintática específica. É mais apropriado para detecção de vários padrões sintáticos entre duas palavras relacionadas, como por exemplo “bater”, “porta”. As palavras que aparecem entre os dois termos variam e a distância entre eles também não é constante. Mas a regularidade nos tipos de ocorrências permite determinar que “bater” é o verbo utilizado neste tipo de situação, e não “golpear” ou “topar”.

O método funciona com base na contagem de palavras vizinhas àquelas que estão sendo testadas. Calcula-se a média e a variância do deslocamento

entre as duas palavras em um *cópus*. Tomemos o seguinte *cópus* de 4 sentenças como exemplo:

- (3.1.) Ela bateu na minha porta.
- (3.2.) Eles bateram na sua porta.
- (3.3) Três mulheres bateram de novo naquela outra porta.
- (3.4) Um homem bateu quinhentas vezes na sua porta.

Contam-se quantas palavras a partir de “bater” existem antes de “porta” em cada uma das quatro frases e calcula-se, a partir de então, a média de deslocamento<sup>18</sup> entre as quatro frases. A média é simplesmente o deslocamento médio.

Já o método de Testagem de Hipóteses se revelou mais adequado para os nossos fins. Tanto pela sua natureza amigável, clara e explicativa quanto pela robustez dos resultados, demonstrou ser bastante pertinente para a detecção de CMs verbais, como demonstraremos a seguir.

### 3.5.1

#### Testagem de hipóteses

Um problema típico de avaliações estatísticas é determinar se algum evento se deve ao acaso ou é motivado. Para tanto, aplica-se a metodologia de Testagem de Hipóteses. É preciso formular uma hipótese nula que apresenta a relação entre duas palavras que não formam uma CM. Para tanto, assumimos que as duas palavras são completamente independentes. A probabilidade de ocorrerem é dada por:

$$H_0: P(w_1 w_2) = P(w_1)P(w_2)$$

Onde H = hipótese, P = probabilidade, w = palavra

---

<sup>18</sup> A média de deslocamento entre *bater* e *porta* é calculada assim:  $1 \frac{(3+3+5+5)}{4} = 4.0$

Dentre os métodos de Testagem de Hipótese, o *Logaritmo de Verossimilhança* tem sido um dos mais utilizados. Ele tem por objetivo detectar se um bigrama é uma CM ou uma co-ocorrência casual na língua. Esse tipo de testagem requer a formulação de dois tipos de hipóteses:

Hipótese 1 (H1): a probabilidade de a ocorrência da primeira palavra de uma seqüência depender da ocorrência da segunda palavra é a mesma do que sua ocorrência independente da segunda palavra;

Hipótese 2, (H2): a probabilidade de a ocorrência da primeira palavra de uma seqüência depender da ocorrência da segunda palavra não é a mesma do que sua ocorrência independente da segunda palavra.

Essas duas hipóteses estão formalizadas abaixo:

$$H1: P(w_1 | w_2) = P(w_1 | \neg w_2)$$

$$H2: P(w_1 | w_2) \neq P(w_1 | \neg w_2)$$

Onde H = hipótese, P = probabilidade, w = palavra

Por exemplo, assumindo que a expressão *fazer sucesso* seja uma CM,

$$H2: P(\text{fazer} | \text{sucesso}) \neq P(\text{fazer} | \neg \text{sucesso})$$

espera-se que a hipótese de independência

$$H1: P(\text{fazer} | \text{sucesso}) = P(\text{fazer} | \neg \text{sucesso})$$

seja falsa. Portanto, o método avalia a probabilidade de H2 ocorrer em detrimento de H1.

A aplicação do Logaritmo de Verossimilhança, disponibilizado através do pacote estatístico NSP (Banerjee & Pedersen, 2003), foi viabilizado, nesse projeto, através de um programa, feito em linguagem Java™, que recebe como entrada o *cópus* e fornece como resultado a lista de todas as co-ocorrências do tipo V+(det)+ N (Nogueira, 2004). Só a partir de então, aplica-se o teste estatístico e é estabelecida a lista das candidatas a CMs que, posteriormente, são ordenadas por frequência. Esse teste será aplicado aos dez verbos com o padrão procurado (em negrito na tabela 3) .

### 3.5.1.1

#### O teste e a avaliação dos resultados

De forma clara, o teste é capaz de trazer à tona combinações do tipo V+(det)+N consideradas pela literatura representacionista como distintivas se comparadas a uma combinação sintática tradicional. Embora estejamos evitando este tipo de consideração, o resultado a que chegamos já se mostra produtivo por demonstrar que um método semanticamente cego é capaz não só de extrair as CMs do cópuz, como também de identificar aquelas mais frequentes. Uma outra virtude do método é a sua eficácia na detecção de determinante, um aspecto relevante no estudo das CMs do tipo V+SN (cf. Garrão, 2001, Vale, 2002). Além disso, o cópuz também se revelou bastante profícuo para a tarefa proposta. A variedade das seções, colunas e matérias do jornal garantiram um resultado abrangente.

Sob uma perspectiva quantitativa, portanto, o método se revelou satisfatório. Em outras palavras, dentre as 1000 candidatas a CMs apontadas pelo método, apenas 128 foram consideradas ruído. Um acerto de 87,2%<sup>19</sup>.

As listas das 1000 CMs detectadas, divididas em 100 ocorrências por verbo, seguem no final desta seção. As “pseudo-CMs” extraídas pelo método, ou seja, os “deslizes” por ele cometido (12,8%), estão indicados na lista da seguinte forma:

- i) Erro de avaliação estrutural. Este tipo de erro pode ter sido cometido pelo método por duas razões principais:

- ❖ (ETQ)

Em função da etiquetagem equivocada no cópuz: por exemplo, algumas combinações nominais compostas, como *reforma agrária*, são consideradas como tais. Isso auxilia a eficácia do método, que detecta satisfatoriamente a CM *fazer a reforma agrária*. Em contrapartida, há palavras que estão erradamente etiquetadas, como por exemplo *livre* (em *Romário recebeu livre dentro da área*), que está anotada no cópuz como nome (SN), talvez em função de sua

<sup>19</sup> Não incluímos nessa avaliação quantitativa o teste 10 (VER). O verbo CRIAR, por ser o 11º mais frequente para o padrão de CM procurado, foi alçado à 10ª colocação.

posição nominal, quando seria adjetivo (Adj). A identificação de CMs como *ganhar líquido* e *ganhar real* também é consequência de erros de etiquetagem (“ganho” foi identificado como primeira pessoa do singular do verbo, quando, na verdade, trata-se de um nome).

❖ (JAN)

Em função de o método ter considerado uma janela sintática menor do que a expressão representa: *ter um papel*, por exemplo, foi detectado pelo método como uma CM do padrão procurado quando, na verdade, sua estrutura vai além de V+(det)+N. Seria V+(det)+N+(Adv)+Adj: *Ter um papel muito importante*. O mesmo acontece com *fazer um trabalho*, (*Paulo Roberto, ainda sem forma física, faz um trabalho apenas burocrático*), e *ganhar linhas* (*A carroceria ganhou linhas mais atuais*). Esse desliz é consequência do formalismo utilizado para restringir o escopo da busca por uma estrutura V+(det) +N que antecederesse um advérbio. Uma outra instância de ruído foi gerada em função de estruturas sintáticas invertidas, deslocadas ou descontínuas: um exemplo é *ganhar este ano* (*a campanha ganhou este ano novo fôlego diante da participação dos jornais*) que foi detectada ao invés de *ganhar novo fôlego*. Um outro exemplo é *ter acesso* (“*O mundo dos fanzines tem algo de sociedade secreta, na qual apenas alguns iniciados têm acesso.*”) Nesse último caso, o ruído foi gerado em virtude da restrição de pontuação, que, por outro lado, foi altamente econômica e compatível com o que queria ser alcançado, como demonstramos anteriormente nesta seção (p.47).

ii) Outros ruídos foram atribuídos exclusivamente ao cópulus:

❖ (DAT)

CMs claramente datadas: São elas: *criar a URV, usar a URV, tomar AZT*.

❖ (MET)

Metatexto do cópulus: encontrado exclusivamente no teste 10 (*ver resultados, ver tabela, ver fotos, etc*), que foi excluído pelo teor metalingüístico de muitas de suas CMs.

Há outros dois tipos de interferência na detecção de CMs que não foram considerados propriamente ruídos. São eles:

❖ (COE)

Recursos coesivos, como a utilização de anáfora: alguns exemplos são *fazer a denúncia*, *dar a notícia*, *ter a doença*. Na verdade, não consideramos esses casos como pseudo-CMs. Ressaltamos apenas que a presença do determinante (na maioria dos casos, um artigo definido) não foi detectada em todos os casos em que a CM ocorre. O mesmo não acontece com CMs como *fazer o contrário*, *dar as caras*, *perder a virgindade*, *usar o bom senso*, dentre outras, cujas estruturas só foram detectadas dessa forma.

❖ (MAN)

Omissões de artigo (tanto definido quanto indefinido), características de manchetes de jornal, como *Presidente da Shell **deixa cargo** amanhã*. Embora a descrição do *cópus* argumente ter eliminado todas as manchetes, isso parece não proceder em todos os casos. Esse tipo de recurso também não foi avaliado como pseudo-CM.

Finalmente, é importante atentar para o fato de que algumas CMs têm mais de uma estrutura transitiva. É o caso, dentre outras CMs, de *fazer campanha*: V+N (*Até agora, apenas Munhoz tem **feito campanha***.) e V+N+Prep+(det)+N, (que foi detectada em sua estrutura deslocada em *O secretário Antonio Felix Domingues **faria campanha**, em Goiás e Tocantins, para o ex-governador Orestes Quércia nas prévias do PMDB*). Como há a possibilidade de a expressão desempenhar um papel intransitivo, não foi identificada na listagem como uma pseudo-CM.

Em suma, a grande vantagem deste método está no seu teor preditivo. Através dele, podemos constatar preferências de usos das expressões presentes no *cópus*. A seguir, apontamos exemplos que ilustram o grau de eficiência do método. Incluímos também alguns exemplos de ruído, ressaltando a necessidade de uma revisão humana:

No teste 1 (“FAZER”):

A CM *fazer parte* foi apresentada pelo *cópus* também na forma intransitiva, embora, intuitivamente, talvez pudéssemos afirmar que a expressão exige um complemento transitivo indireto linguisticamente explícito.

- *Ciúme dosado **faz parte.***
- *O medo **faz parte.***
- *Perder gol **faz parte.***
- *Lógico que tem um lado burocrático, mas **faz parte;***

Por essa razão, foi considerada uma CM.

Já *fazer comentários*, embora também tenha sido avaliada como uma CM (*Iasser Arafat não foi encontrado para **fazer comentários***), parece estar atrelada a uma janela estrutural maior encabeçada pelo ADV *não* ou alguma consideração de teor negativo como a preposição *sem*, o verbo *evitar* ou a expressão verbal *não querer* :

- *Iberê Camargo as leu, mas não **fez comentários.***
- *Roseana também não **faz comentários.***
- *Itamar e June trocaram presentes, mas não **fizeram comentários.***
- *Ao sair, Cerqueira preferiu não **fazer comentários.***
- *Tim ignorou a regra de receber o troféu e deixar o palco **sem fazer comentários.***
- *Luxemburgo evitou **fazer comentários.***

No teste 2 (“TER”):

A CM *ter uma relação* também parece ter um uso intransitivo, além do seu teor transitivo indireto tradicional.

• *Quando eu estava sozinha, sentia falta de **ter uma relação, não de trepar por trepar.***

É possível detectar, também, um padrão de ocorrência característico de CMs encabeçadas pelo verbo *ter*. V+N (*ter caráter, ter valor*) e V+(det)+N+(adv)+Adj (*ter um caráter Adj*) (*ter um valor Adj*):

- *Chicago é uma cidade que tem alma, **tem caráter.***

- *Foram tenistas que não precisavam de «personalidade», porque **tinham caráter**.*
- *Esta punição ao tenente-coronel **tem um caráter exemplar** ?*
- *Portanto, considero que isso **tem um caráter institucional**.*
- *Ser negro é ser a paz, **ter valor**.*
- *Os camaroneses **tiveram valor**, os brasileiros, classe», afirma o jornal .*
- *Um aluno desse **tem um valor incrível**.*
- *No final, quando acabo, sinto que eles podem **ter um valor alegórico, simbólico**.*

O mesmo ocorre com *ter validade*. Já *ter defeito* não segue o mesmo padrão. Aparece também intercalado do artigo indefinido com uma estrutura intransitiva:

- *Os figos só **têm um defeito**: não são recomendados para quem está de dieta .*
- *“Quando um filme **tem um defeito**, é condenado como obra mal realizada” .*

No Teste 3 (“DAR”):

As expressões *dar banho*, *dar medo*, *dar voto*, por exemplo, foram consideradas CMs pela sua intransitividade em algumas ocorrências.

- *Troco fraldas, **dou banho**, e depois me tranco lá dentro por pelo menos três horas por dia .*
- *Você vai ter que **dar banho**, talvez trocar fraldas .*
- *Mulher independente **dá medo**.*
- *O Brasil **dá medo**.*
- *É, ferrovia nunca **deu voto**.*
- *Quer aquilo que **dá voto**.*

Embora inusitada, a expressão *dar outras providências* é freqüente no *cópus* em virtude de seu teor jurídico:

- *Lei 6.766, de 19.12.79 - Dispõe sobre o parcelamento do solo urbano e **dá outras providências** (CCLCV 352, Lex 1979/1.008, Bo)l .*

- 215: 2 a. LC 73, de 10.2.93 - *Institui a Lei Orgânica da Advocacia-Geral da União e dá outras providências* (Lex 1993/82, RF 321 / 420) :

- Lei 8.021, de 12.4.90 - *Dispõe sobre a identificação dos contribuintes para fins fiscais, e dá outras providências* (Lex 1990/518, RDA 179 / 348):

#### No Teste 4 (“PERDER”):

De uma forma ampla, o teste não gerou ruídos em relação a CMs encabeçadas pelo verbo “perder”. Isso demonstra não só que o formalismo utilizado para detectar as CMs tem um bom poder de abrangência mas também que o verbo “perder” é vocacionado a encabeçar CMs do padrão procurado.

Pode-se questionar a identificação de “perder móveis” como uma CM freqüente; o que pode ser explicado pelo teor jornalístico do córpus.

- *Vizinho de Deosdeti, o bombeiro **perdeu móveis**, mas retornou ontem à rua onde mora para se assegurar de que não roubaram o que lhe restou.*

Notamos um padrão enumerativo com esta expressão:

- *Na favela da Santa Cruz 2, 18 famílias **perderam móveis**, roupas e mantimentos .*

- *Pelo menos cem famílias de classe média na Vila Ema foram atingidas e **perderam móveis**, eletrodomésticos e mantimentos .*

- *Os moradores **perderam móveis**, objetos e até carros .*

#### No teste 5 (“USAR”):

Aqui houve um ruído específico gerado não pelo método estatístico implementado, mas pela época em que o córpus foi compilado. As CMs *usar a URV* e *usar URV*:

- *A Receita Federal não quer **usar a URV**.*

- *A outra é sobre como será o relacionamento com os fornecedores, já que uns **usarão URV** e outros não .*

• *Com isso as empresas podem **usar a URV**, mas não são obrigadas a adotar esse procedimento», explicou Roberto Padovani, da Secretaria de Política Econômica do Ministério da Fazenda .*

No Teste 6 (“RECEBER”):

A expressão “receber tratamento”, por exemplo, ocorre frequentemente com padrão intransitivo, com um uso bastante definido:

- *Para o ministério, estudantes com visão inferior à normal devem **receber tratamento** .*
- *A foca foi levada para o zôo, em São Cristóvão, onde **receberá tratamento** .*
- *Até as 6h de sábado, ele permanecia em uma maca **recebendo tratamento** .*

Já quando o padrão é “receber +um+ tratamento”, a janela sintática parece ser “receber +um+ tratamento+ (Adv)+Adj”, o que aumenta a possibilidade de aplicações para seus usos:

- *E, a de uma clínica de emagrecimento, a Proforma, a cantora de 19 anos **recebeu um tratamento gratuito** .*
- *Sendas já adiantou que os fornecedores alinhados com a promoção **receberão um tratamento vip** .*
- *Os candidatos à Presidência da República e a vice **recebem um tratamento ainda mais vip**.*
- *E vai **receber um tratamento diferenciado** .*
- *Molina afirmou que o leite in natura distribuído para a população não **recebe um tratamento adequado** .*
- *A grande inovação é o tecido de fibras sintéticas que **recebe um tratamento antichamas** .*

De uma forma geral, o verbo “receber” gerou poucos ruídos (3 pseudo-CMs: *receber um tratamento, receber livre, receber este ano*).

Teste 7 (“DEIXAR”):

O verbo “deixar” também demonstra ser vocacionado a encabeçar CMs do padrão procurado. Houve apenas 5 ruídos. Dentre eles, um merece destaque: “deixar patente”. A ambigüidade do termo e a posição que exercia na estrutura causaram uma etiquetagem equivocada pelo córpus. “Patente”, nesse caso, é adjetivo e não um substantivo:

- *O que há de alegórico não propriamente no filme, mas na sua difusão televisual, é que ele **deixa patente** (pela inserção dos peixinhos coloridos na cena monocromática) e ao mesmo tempo tematiza no enredo a lacuna que é o preto-e-branco televisual .*

Por outro lado, não houve nenhuma ocorrência no córpus da possível CM “deixar (det) patente” (“deixar o cargo militar”).

“Deixar pistas” e “deixar rastros”, por sua vez, foram consideradas pseudo-CMs (JAN) por só terem sido detectadas precedidas pelo advérbio *não* ou pela preposição *sem* :

- *«A explicação é que o garoto resolve assaltar o cliente e, para não **deixar pistas**, acaba matando», diz Mott .*
- *O jogo consiste em matar pessoas violentamente e não **deixar pistas** .*
- *Há cinco anos, um Cessna 720, com capacidade para quatro passageiros, foi levado do aeroporto sem **deixar pistas** .*
- *À moda de um bom criminoso hitchcockiano: sem **deixar rastros** .*
- *que a ex-senhora Mick deixou algumas jóias no cofre do hotel do amigo, o Morgans Hotel, em NY, e elas desapareceram sem **deixar rastros** .*
- *Tinha, pois, charme suficiente para cumprir breve passagem no Santos, aquele Santos de Pelé e cia. Tão breve que não **deixou rastros** .*

No Teste 8 (“TOMAR”):

A expressão “tomar conta”, por exemplo não foi considerada uma pseudo-CM, uma vez que aparece em estruturas intransitivas com um teor específico:

- *As pessoas pensavam que, sem as crenças religiosas tradicionais, o cinismo ia **tomar conta**.*

- *Durante os primeiros meses, pratica Tai Chi com entusiasmo mas aos poucos o cansaço vai **tomando conta**, e resolve desistir .*
- *Havíamos atracado num pequeno desembarcadouro, sede de um projeto governamental abandonado, onde a margem arenosa se estendia por mais de 30 metros antes da floresta **tomar conta**.*
- *Se o ator não souber parar, a vida profissional vai sempre **tomar conta**.*
- *Aí vem a mente, bocejando, e **toma conta**.*

Por outro lado, “tomar o AZT”, “tomar AZT” foram consideradas ruídos pela sua frequência ter sido gerada em função da época em que o córpus foi coletado:

- *Um remédio usado contra herpes, o acyclovir, aumentou em 44 % a sobrevivência de pacientes de Aids que também **tomavam o AZT**, segundo estudo publicado ontem na revista «Annals of Internal Medicine» .*
- *Dois meses atrás, a Newsweek publicou uma notícia afirmando que ele estava **tomando AZT** e que tinha visitado uma clínica no Quênia para tratamento .*

No Teste 9 (“GANHAR”):

Tanto a expressão “ganhar vida” quanto “ganhar a vida” foram consideradas CMs. A primeira, em razão de sempre desempenhar uma função intransitiva no córpus:

- *Ela também divide o palco com objetos que **ganham vida**.*
- *Espero que a experiência tome curso diferente e **ganhe vida**.*

Já a segunda, por frequentemente desempenhar uma função intransitiva:

- *Fazer filmes é uma profissão, um jeito de **ganhar a vida**.*
- *“Eu estou **ganhando a vida!** ”, me diz o cu-a-jato- com um jeito meio sindicalista .*
- *“Temos que **ganhar a vida**. ”; além de seu papel transitivo indireto*

tradicional:

- *Quatro vezes campeão mundial de F-1, Alain Prost **ganha a vida**, hoje em dia, como comentarista da TF1, canal de TV da França .*

Um ruído provocado pelo verbo “ganhar”, foi gerado pela ambigüidade do termo “ganho” (substantivo), detectado como forma flexionada do verbo. Houve, conseqüentemente, a identificação de pseudo-CMs como “ganhar real”, “ganhar líquido” e “ganhar financeiro”.

- **Ganho líquido:** 4,94 % .
- *Os analistas recomendam a retirada dos recursos somente dez dias depois do depósito, no mínimo, para garantir **ganho real**.*
- *Com a entrada da URV, os supermercados deixaram de ter **ganho financeiro**, diz Diniz .*

No Teste 10 (“VER”):

Conforme já argumentamos, as CMs encabeçadas pelo verbo “ver” foram excluídas em virtude de sua freqüente função metatextual. “Ver quadro”, por exemplo, ocorre 150 vezes no cópús. “Ver texto”, por sua vez, ocorre 65 vezes. Claramente, a sua posição entre os 10 verbos mais recorrentes com o padrão procurado está comprometida pelo cópús.

No Teste 11 (“CRIAR”):

A expressão *criar chances* foi considerada uma CM pela sua função freqüentemente intransitiva em parágrafos referentes às seções de Esportes.

- *Na prorrogação, os dois times **criaram chances**, mas ninguém marcou .*
- *O Corinthians teve maior posse de bola, atacou, **criou chances** e chutou mais contra o gol de Velloso, mas não conseguiu ir além do empate por 1 a 1, que deu o bicampeonato ao Palmeiras, ontem no Pacaembu .*
- *O trio, porém, não conseguiu **criar chances** .*

O mesmo foi verificado em relação a ***criar um ambiente*** e ***criar um cenário***:

- *Então foi **criado um ambiente**.*
- *No fundo, você está **criando um cenário**, criando uma falsa idéia .*

\* \* \*

Em suma, o que consideramos especialmente relevante nesta abordagem com base em *córpus*, é que não fazemos conjecturas daquilo que ocorre e não ocorre em uma língua, pois, como já demonstramos no capítulo 2, uma perspectiva exclusivamente intuitiva pode ser muitas vezes contra-argumentada pelo *córpus*. Nosso olhar eminentemente empírico é capaz de detectar preferências de usos ao invés de intuir aquilo que pode ou não ocorrer em um *córpus*, com base em testes de aceitabilidade, também já questionados no capítulo 2.

As listas referentes aos TESTES 1 a 11 estão apresentadas em ordem decrescente de frequência de cada verbo com o padrão V+(det)+N procurado. As pseudo-CMs estão indicadas por parênteses, seguidas pela sigla do ruído.

Na verdade, a listagem de cada verbo testado segue até que se chegue a CMs menos frequentes (com uma única ocorrência). Optamos por expor as 100 CMs mais recorrentes encabeçadas por cada um dos dez verbos por razões metodológicas.

Nas 10 listas, cada CM é acompanhada pelo seu ranqueamento e frequência, respectivamente. Tomemos como exemplo o primeiro bigrama do TESTE 1, [*fazer campanha*]: o número que segue imediatamente a CM (1) diz respeito à sua posição no *córpus* em relação às outras CMs encabeçadas pelo mesmo verbo. Já o segundo número (117) se refere ao número de ocorrências da CM [*fazer campanha*] e de suas diversas flexões no *córpus*. Note-se que as CMs estão agrupadas na forma canônica do verbo, mas são identificadas em todas as suas formas flexionadas.

### **Abreviaturas:**

- (ETQ): etiquetagem equivocada no *córpus*
- (JAN): janela sintática menor do que a expressão representa
- (DAT): CMs claramente datadas
- (MET): metatexto do *córpus*
- (COE): Recursos coesivos
- (MAN): manchetes de jornal

**Teste 1: FAZER +(DET)+N**

- fazer campanha,1,117  
 fazer parte,2,106,  
 fazer sucesso,3,96  
 fazer sentido,4,94  
 fazer compras,5,73  
 fazer falta,6,57  
 fazer perguntas,7,51  
 fazer alguma coisa,8,44  
 fazer política,9,44  
 fazer as contas,10,42(COE)  
 fazer gols,11,40  
 (fazer mal),12,39 (JAN)  
 fazer o gol,13,38 (COE)  
 fazer greve,14,36  
 fazer sexo,15,36  
 fazer filmes,16,35  
 fazer ginástica,17,34  
 fazer cinema,18,34  
 fazer amor,19,32  
 fazer um filme,20,30  
 fazer a conversão,21,30 (COE)  
 fazer comício,22,30  
 (fazer muito=tempo),23,29 (JAN)  
 fazer diferença,24,29  
 fazer água,25,28  
 fazer um teste,26,27  
 fazer o pedido,27,26  
 fazer as coisas,28,26 (COE)  
 fazer exercícios,29,25  
 fazer a festa,30,25  
 fazer o teste,31,24 (COE)  
 fazer boca-de-urna,32,24  
 fazer a revisão,33,24 (COE)  
 (fazer coisas),34,24 (JAN)  
 fazer shows,35,24  
 fazer tempo,36,23  
 fazer a mesma coisa,37,23 (COE)  
 (fazer uma coisa),38,23 (JAN)  
 fazer teatro,39,22  
 fazer o filme,40,21 (COE)  
 fazer exames,41,21  
 fazer concessões,42,21  
 fazer comentários,43,21  
 fazer escola,44,20  
 fazer teste,45,20  
 fazer outra=coisa,46,20 (COE)  
 fazer palestra,47,20  
 fazer gol,48,20  
 fazer testes,49,20  
 fazer contas,50,19  
 fazer dinheiro,51,19  
 fazer a denúncia,52,19 (COE)  
 fazer propaganda,53,19  
 fazer a sua parte,54,19,1  
 fazer carreata,55,19,1  
 fazer acordo,56,18,1  
 fazer uma ressalva,57,18,1  
 fazer música,58,18,1  
 fazer a diferença,59,18,1  
 fazer milagres,60,18,1  
 fazer coro,61,18,1  
 fazer barulho,62,18  
 fazer críticas,63,18  
 fazer a prova,64,17(COE)  
 fazer o exame,65,17(COE)  
 fazer ameaças,66,17  
 fazer um acordo,67,17  
 fazer justiça,68,16  
 fazer doações,69,16  
 fazer as malas,70,16  
 fazer a defesa,71,15 (COE)  
 fazer sua estréia,72,15 (COE)  
 fazer o trabalho,73,15 (COE)  
 fazer show,74,15  
 fazer assembléia,75,15  
 fazer acordos,76,15  
 fazer dieta,77,14  
 fazer oposição,78,14  
 fazer a reforma=agrária,79,14 (COE)  
 fazer reféns,80,14  
 fazer amigos,81,14  
 fazer uma curva,82,14  
 fazer pesquisas,83,13  
 fazer a barba,84,13

fazer seu sucessor,85,13 (COE)  
fazer previsões,86,13  
fazer reservas,87,13  
fazer o pagamento,88,13 (COE)  
fazer essas declarações,89,13 (COE)  
fazer a coisa,90,12 (COE)  
fazer sol,91,12  
(fazer um trabalho),92,12 (JAN)  
fazer promoções,93,12  
(fazer tanto sucesso),94,12 (JAN)  
fazer mudanças,95,12  
fazer aborto,96,12  
fazer análise,97,12  
fazer mais gols,98,11 (COE)  
fazer alianças,99,11  
fazer turismo,100,11  
fazer muito sucesso,101,11  
fazer uma pesquisa,102,11  
fazer o serviço,103,11(COE)  
fazer um discurso,104,11  
fazer o projeto,105,11(COE)  
fazer festa,106,11

**Teste 2: TER+(DET)+N**

ter razão,1,161  
 ter problemas,2,135  
 (ter direito),3,119 (JAN)  
 ter medo,4,114  
 ter uma idéia,5,107  
 (ter acesso),6,94 (JAN)  
 ter filhos,7,92  
 (ter certeza),8,75 (JAN)  
 ter dinheiro,9,68  
 ter sucesso,10,67  
 ter dúvidas,11,61  
 ter início,12,60  
 ter tempo,13,58  
 ter alta,14,54  
 ter jeito,15,53  
 ter paciência,16,50  
 ter sorte,17,49  
 ter um filho,18,48  
 ter problema,19,45  
 (ter nome),20,40 (JAN)  
 ter importância,21,40  
 ter dificuldades,22,32  
 ter limites,23,32  
 ter experiência,24,30  
 ter valor,25,30  
 ter dúvida,26,30  
 ter validade,27,29  
 ter candidato,28,29  
 ter uma visão,29,29  
 ter notícia,30,29  
 ter fim,31,28  
 ter condições,32,28  
 ter êxito,33,27  
 ter preço,34,26  
 ter fundamento,35,25  
 ter chance,36,25  
 ter coragem,37,24  
 ter força,38,24  
 ter uma filha,39,24  
 ter prejuízo,40,24  
 ter solução,41,23  
 ter telefone,42,23  
 ter culpa,43,23  
 ter chances,44,22  
 ter pressa,45,22  
 ter qualidade,46,22  
 (ter interesse),47,20 (JAN)  
 ter fôlego,48,20  
 ter queda,49,20  
 ter espaço,50,20  
 ter maioria,51,20  
 ter provas,52,20  
 ter escrúpulos,53,20  
 (ter conhecimento),54,19 (JAN)  
 ter uma fazenda,55,19  
 ter idéias,56,19  
 ter talento,57,19  
 ter poder,58,19  
 ter a mesma opinião,59,18(COE)  
 ter futuro,60,18  
 ter a doença,61,18(COE)  
 ter esperança,62,18  
 ter mais chances,63,17  
 ter uma explicação,64,17  
 (ter a impressão),65,17 (JAN)  
 ter defeitos,66,17  
 ter lugar,67,17  
 ter cura,68,17  
 ter prioridade,69,17  
 ter muito=tempo,70,16  
 ter muito dinheiro,71,16  
 ter música,72,16  
 ter saída,73,16  
 ter efeito,74,16  
 ter credibilidade,75,16  
 ter vez,76,16  
 (ter um caráter),77,15 (JAN)  
 ter dono,78,15  
 (ter preços),79,15(JAN)  
 (ter seu preço),80,15 (JAN)  
 ter história,81,15  
 (ter gente),82,15(JAN)

ter um problema,83,15  
ter uma relação,84,14  
ter memória,85,14  
ter lucro,86,14  
ter mais votos,87,14  
ter aumento,88,14  
ter continuidade,89,14  
(ter recursos),90,14 (JAN)  
ter câncer,91,14  
ter amigos,92,14  
ter liberdade,93,14  
(ter vontade),94,14 (JAN)  
ter inflação,95,13  
ter transmissão,96,13  
(ter um papel),97,13 (JAN)  
(ter um desempenho),98,13 (JAN)  
ter a bola,99,13  
ter resultados,100,13  
ter preferência,101,13  
ter um estilo,102,13  
(ter idéia),103,13 (JAN)  
ter um jogo,104,13  
(ter coisa),105,12(JAN)  
ter fé,106,12  
ter computador,107,12  
ter erro,108,12  
ter efeitos,109,12 (JAN)  
ter caráter,110,12  
ter água,111,12  
ter um impacto,112,12  
ter um efeito,113,12  
ter cabimento, 114,12  
ter piscina,115,12  
ter um preço,116,12  
(ter mais dinheiro),117,11(JAN)  
ter um custo,118,11  
ter uma reunião,119,11

**Teste 3: DAR+(DET)+N**

dar entrevistas,1,115  
 dar entrevista,2,67  
 dar resultado,3,46  
 (dar a volta),4,42 (JAN)  
 (dar declarações),5,42 (JAN)  
 dar lucro,6,38  
 dar sorte,7,37  
 dar aulas,8,37  
 dar um exemplo,9,36  
 dar autógrafos,10,35  
 dar início,11,29  
 (dar conta),12,28 (JAN)  
 (dar explicações),13,27 (JAN)  
 dar tempo,14,26  
 dar o troco,15,25  
 dar dinheiro,16,24  
 dar resultados,17,23  
 dar prejuízo,18,22  
 dar um tempo,19,22  
 dar um salto,20,22  
 dar trabalho,21,22  
 dar o tom,22,21  
 dar um passo,23,20  
 dar as caras,24,20  
 dar risada,25,19  
 dar detalhes,26,19  
 (dar o nome),27,18 (JAN)  
 dar o exemplo,28,18 (COE)  
 dar força,29,17  
 dar apoio,30,17  
 dar espetáculo,31,16  
 (dar origem),32,16 (JAN)  
 (dar momento),33,16 (ETQ)  
 dar a resposta,34,15 (COE)  
 (dar comida),35,14(JAN)  
 dar atenção,36,14  
 (dar sustentação),37,13 (JAN)  
 dar medo,38,13  
 dar azar,39,13  
 dar descontos,40,13  
 (dar continuidade),41,13 (JAN)  
 dar frutos,42,13  
 (dar lugar),43,12 (JAN)  
 (dar informações),44,12(JAN)  
 dar aula,45,12  
 dar as cartas,46,12  
 dar muito trabalho,47,12  
 dar uma entrevista,48,11  
 dar vexame,49,11  
 (dar uma idéia),50,11 (JAN)  
 (dar espaço),51,11 (JAN)  
 dar as mãos,52,11  
 dar nomes,53,11  
 (dar bola),54,11 (JAN)  
 dar outras providências,55,11  
 dar um jeito,56,11  
 dar porrada,57,10  
 (dar ouvidos),58,10 (JAN)  
 dar palestras,59,10  
 (dar opinião),60,10 (JAN)  
 (dar razão),61,9 (JAN)  
 dar a partida,62,9  
 dar prazer,63,9  
 dar respostas,64,9  
 dar voto,65,9  
 (dar a impressão),66,9 (JAN)  
 dar conselhos,67,9  
 dar um conselho,68,9  
 dar retorno,69,9  
 dar uma força,70,9  
 (dar a notícia),71,9 (COE)  
 dar entrada,72,8  
 dar exemplos,73,8  
 dar palestra,74,8  
 dar orientações,75,8  
 dar problema,76,8  
 dar uma guinada,77,8  
 (dar a vitória),78,8 (JAN)  
 (dar um carro),79,7 (JAN)  
 (dar alegria),80,7 (JAN)  
 dar palpite,81,7  
 dar pena,82,7  
 dar bandeira,83,7  
 dar dicas,84,7  
 dar futuro,85,7

dar manchete,86,7  
 dar seu parecer,87,7(COE)  
 dar uma festa,88,7  
 dar chance,89,7  
 (dar sinais),90,7 (JAN)  
 dar tiros,91,7  
 dar pênalti,92,7  
 dar banho,93,7  
 dar pé,94,7  
 dar samba,95,7  
 dar problemas,96,7  
 dar resposta,97,6  
 dar as informações,98,6 (COE)  
 (dar sequência),99,6 (JAN)  
 dar mais trabalho,100,6  
 (dar esclarecimentos),101,6 (JAN)  
 (dar valor),102,6 (JAN)  
 dar prestígio,103,6  
 (dar prosseguimento),104,6 (JAN)  
 dar voltas,105,6  
 dar troco,106,6  
 dar votos,107,6  
 (dar destaque),108,6 (JAN)  
 dar um show,109,6  
 dar um toque,110,6  
 (dar membro),111,6 (ETQ)  
 dar show,112,6  
 dar a receita,113,6  
 (dar espaços),114,5 (JAN)  
 dar um abraço,115,5  
 dar o dinheiro,116,5  
 (dar um caráter),117,5 (JAN)  
 dar jeito,118,5  
 dar os nomes,119,5 (COE)  
 dar confiança,120,5  
 dar alguns exemplos,121,5  
 dar o recado,122,5 (COE)  
 dar sua opinião,123,5 (COE)  
 dar queixa,124,5  
 dar a informação,125,5 (COE)  
 dar sugestões,126,5  
 dar um parecer,127,5  
 dar fama,128,5  
 dar licença,129,5  
 (dar cobertura),130,5 (JAN)  
 dar esmola,131,5

**Teste 4: PERDER+(DET)+N**

- perder tempo,1,54  
 perder a eleição,2,46  
 perder o controle,3,44  
 perder dinheiro,4,41  
 perder o emprego,5,38  
 perder peso,6,28  
 perder a bola,7,25  
 perder o mandato,8,23  
 perder força,9,20  
 perder a direção,10,19  
 perder o fôlego,11,17  
 perder a cabeça,12,15  
 perder a validade,13,14  
 perder o sentido,14,14 (COE)  
 perder as eleições,15,13  
 perder o título,16,13 (COE)  
 perder espaço,17,13  
 perder o rumo,18,13  
 perder a pose,19,12  
 perder o ritmo,20,11  
 perder o registro,21,11  
 perder a mãe,22,11  
 perder o jogo,23,11(COE)  
 perder a graça,24,10  
 perder um pênalti,25,10  
 perder a paciência,26,10  
 perder votos,27,10  
 perder valor,28,10  
 perder a razão,29,9  
 perder a copa,30,9  
 perder a virgindade,31,9  
 perder a oportunidade,32,9  
 perder o equilíbrio,33,9  
 perder o cargo,34,9  
 perder sentido,35,9  
 perder terreno,36,8  
 perder o interesse,37,8 (COE)  
 perder pontos,38,8  
 perder o pai,39,8  
 perder seus  
 empregos,40,8(COE)  
 perder a calma,41,8  
 perder a força,42,8 (COE)  
 perder fôlego,43,8  
 perder competitividade,44,8  
 perder a esperança,45,8  
 perder validade,46,8  
 perder a guerra,47,7 (COE)  
 perder o pique,48,7  
 perder emprego,49,7  
 perder credibilidade,50,7  
 perder a majestade,51,7  
 perder móveis,52,7  
 perder mercado,53,7  
 perder o medo,54,6  
 perder o bom=humor,55,6  
 perder gols,56,6  
 perder o costume,57,6  
 perder o valor,58,6 (COE)  
 perder a qualidade,59,6  
 perder a mulher,60,6  
 perder a memória,61,6  
 perder a partida,62,6(COE)  
 perder as esperanças,63,6  
 perder a vaga,64,5  
 perder suas casas,65,5 (COE)  
 perder a posição,66,5  
 perder o lugar,67,5  
 perder seu mandato,68,5(COE)  
 perder o gol,69,5  
 perder a chance,70,5(COE)  
 perder eleições,71,5  
 perder os dedos,72,5  
 perder o marido,73,5  
 perder clientes,74,5  
 perder eficácia,75,5  
 perder a voz,76,5  
 perder o rebolado,77,5  
 perder eficiência,78,5  
 perder muito=tempo,79,5,  
 perder a ação,80,5 (COE)  
 perder a liderança,81,5  
 perder o filho,82,5 (COE)  
 perder a concentração,83,5  
 perder a causa,84,5 (COE)  
 perder chance,85,4  
 perder seus mandatos,86,4,(COE)

perder importância,87,4  
perder o foco,88,4  
perder várias chances,89,4  
perder pênalti,90,4(COE)  
perder um jogador,91,4  
perder o apetite,92,4  
perder a conta,93,4  
perder a vez,94,4  
perder os pontos,95,4  
perder a hora,96,4  
perder os empregos,97,4(COE)  
perder o bonde,98,4  
perder a identidade,99,4  
perder a perna,100,4

**Teste 5: USAR+(DET)+N**

- usar camisinha,1,43  
 usar o cinto,2,36 (COE)  
 usar drogas,3,36  
 usar cinto=de=segurança,4,19  
 usar cinto,5,15  
 usar óculos,6,14  
 (usar a urv),7,12 (DAT)  
 usar a camisinha,8,12 (COE)  
 usar computador,9,12  
 usar preservativo,10,12  
 usar o computador,11,11 (COE)  
 usar o equipamento,12,10 (COE)  
 usar o micro,13,10  
 usar o serviço,14,10 (COE)  
 usar a força,15,9  
 usar terno,16,9  
 usar brinco,17,9  
 usar o cinto=de=segurança,18,8 (COE)  
 usar armas,19,8  
 (usar a máquina),20,8 (JAN)  
 usar calcinha,21,8  
 usar óculos=escuros,22,7  
 usar a cabeça,23,7  
 (usar palavras),24,7 (JAN)  
 usar o sistema,25,7  
 usar cocaína,26,7  
 usar computadores,27,6  
 usar o telefone,28,6  
 usar camiseta,29,6  
 usar a droga,30,6,1(COE)  
 usar capacetes,31,6,1  
 usar chapéu,32,6,1  
 usar o dinheiro,33,6,1(COE)  
 (usar roupas),34,6,1(JAN)  
 usar a rede,35,5,1  
 usar as mãos,36,5,1  
 usar helicópteros,37,5,1  
 usar música,38,5,1  
 usar a piscina,39,5,1  
 usar o programa,40,5,1(COE)  
 usar luvas,41,5,1  
 usar a igreja,42,5,1(COE)  
 usar batom,43,5,1  
 usar água,44,4,1  
 usar armas=de=fogo,45,4,1  
 usar gravata,46,4,1  
 usar os equipamentos,47,4,1  
 usar a violência,48,4,1  
 usar roupa,49,4,1  
 usar a palavra,50,4,1  
 usar meias,51,4,1  
 usar a gráfica,52,4,1 (COE)  
 usar o carro,53,4,1 (COE)  
 usar o dólar,54,4,1  
 usar uniforme,55,3,1  
 usar a técnica,56,3,1(COE)  
 usar uma faca,57,3,1  
 (usar termos),58,3,1(JAN)  
 (usar o jargão),59,3 (JAN)  
 usar máscaras,60,3  
 usar um fone,61,3  
 usar uma camisinha,62,3  
 (usar expressões),63,3(JAN)  
 usar essa palavra,64,3(COE)  
 (usar uma imagem),65,3 (JAN)  
 usar máscara,66,3  
 usar fax,67,3  
 usar crachás,68,3  
 usar o líbero,69,3 (COE)  
 usar véu,70,3  
 usar barba,71,3  
 (usar o livro),72,3 (JAN)  
 usar seu cartão,73,3 (COE)  
 (usar faixas),74,3 (JAN)  
 usar bonés,75,3  
 usar o silêncio,76,3  
 usar cheque,77,3  
 usar maquiagem,78,3  
 usar esse recurso,79,3 (COE)  
 (usar o local),80,3 (JAN)  
 usar o mesmo argumento,81,3 (COE)  
 (usar as importações),82,3 (JAN)  
 (usar o corpo),83,3 (JAN)  
 (usar instrumentos),84,3 (JAN)  
 usar lentes,85,3  
 usar o preservativo,86,3 (COE)

usar ternos,87,3  
usar cueca,88,3  
usar o bom=senso,89,3  
(usar imagens),90,3 (JAN)  
usar pseudônimo,91,3  
usar droga,92,3  
(usar tecnologia),93,3 (JAN)  
usar os programas,94,3 (COE)  
usar passes,95,3  
usar bigode,96,3  
(usar o púlpito),97,3 (JAN)  
usar cores,98,3  
usar crediário,99,3  
usar a tecnologia,100,2  
usar sabão,101,2  
usar pagers,102,2  
(usar urv),103,2 (DAT)  
usar a história,104,2  
usar um computador,105,2

usar uma peruca,106,2  
usar esse instrumento,107,2 (COE)  
usar microcomputadores,108,2  
usar brincos,109,2  
usar disfarces,110,2  
usar carro,111,2  
usar o logotipo,112,2 (COE)  
usar o sol,113,2  
usar as palavras,114,2  
usar o cartão=de=crédito,115,2  
usar líbero,116,2  
(usar amortecedores),117,2 (JAN)  
usar outros índices,118,2 (COE)  
(usar o hotel),119,2 (JAN)  
usar flash,120,2  
(usar a música),121,2 (JAN)  
(usar envelopes),122,2 (JAN)  
(usar a religiosidade),123,2 (JAN)  
(usar bonecos),124,2 (JAN)  
usar as mesmas armas,125,2 (COE)

**Teste 6: RECEBER+(DET)+N**

- receber o dinheiro,1,41(COE)  
 receber alta,2,37  
 receber ameaças,3,27  
 receber a bola,4,26  
 receber dinheiro,5,25  
 receber propinas,6,19  
 receber a notificação,7,16(COE)  
 receber visitas,8,15  
 receber o carro,9,15 (COE)  
 receber salário,10,14  
 receber informações,11,12  
 receber prêmios,12,11  
 receber o benefício,13,10 (COE)  
 receber elogios,14,10  
 receber o produto,15,9 (COE)  
 receber água,16,9  
 receber salários,17,9  
 receber resposta,18,9  
 receber os bônus,19,9 (COE)  
 receber o salário,20,9  
 receber doações,21,8  
 receber denúncias,22,8  
 receber o pagamento,23,8  
 receber o processo,24,7 (COE)  
 receber a notícia,25,7 (COE)  
 receber o prêmio,26,7 (COE)  
 receber seu dinheiro,27,7 (COE)  
 receber um telefonema,28,7  
 receber tratamento,29,7  
 receber a carta,30,7(COE)  
 receber seu salário,31,7 (COE)  
 receber o mesmo tratamento,32,7 (COE)  
 receber ajuda,33,7  
 receber críticas,34,7  
 receber a indenização,35,6 (COE)  
 receber fax,36,6  
 receber uma ameaça,37,6  
 receber armas,38,6  
 receber os salários,39,6  
 receber a diferença,40,6  
 receber a folha,41,6 (COE)
- receber apoio,42,6  
 receber recursos,43,6  
 receber o troféu,44,5  
 (receber este ano),45,5 (JAN)  
 receber a imprensa,46,5  
 receber treinamento,47,5  
 receber o pedido,48,5  
 receber a encomenda,49,5 (COE)  
 receber a denúncia,50,5 (COE)  
 receber pensão,51,5  
 receber benefícios,52,5  
 receber o documento,53,5 (COE)  
 receber comida,54,4  
 receber a restituição,55,4 (COE)  
 receber seus salários,56,4 (COE)  
 receber telefonemas,57,4  
 receber a vacina,58,4 (COE)  
 receber amigos,59,4  
 receber propina,60,4  
 receber os amigos,61,4  
 receber salário=mínimo,62,4  
 (receber um tratamento),63,4 (JAN)  
 receber uma carta,64,4  
 receber oxigênio,65,4  
 receber um prêmio,66,4  
 receber propostas,67,4  
 receber cartas,68,4  
 receber inscrições,69,4  
 (receber livre),70,4 (ETQ)  
 receber dividendos,71,4  
 receber muitas cartas,72,4  
 receber prêmio,73,4 (COE)  
 receber a pensão,74,4 (COE)  
 receber o visto,75,4  
 receber a aposentadoria,76,4  
 receber o troco,77,3  
 receber a revista,78,3 (COE)  
 receber um mínimo,79,3  
 receber esse dinheiro,80,3 (COE)  
 receber reais,81,3  
 receber presentes,82,3  
 receber alimentos,83,3

receber a informação,84,3 (COE)  
receber reclamações,85,3  
receber lançamento,86,3  
receber horas=extras,87,3  
receber tiros,88,3  
receber os cumprimentos,89,3  
receber o telefonema,90,3 (COE)  
receber os dólares,91,3 (COE)  
receber mais passes,92,3 (COE)  
receber esse tratamento,93,3 (COE)  
receber orientação,94,3  
receber uma vaga,95,3  
receber mensagens,96,3  
receber a ligação,97,3(COE)  
receber a conta,98,3 (COE)  
receber os passes,99,3 (COE)  
receber ameaça,100,3  
receber alimentação,101,3  
receber a fatura,102,3 (COE)  
receber juros,103,3

## Teste 7: DEIXAR (DET) N

- deixar o cargo,1,98  
 deixar o país,2,77  
 deixar o governo,3,69  
 deixar o local,4,40  
 deixar filhos,5,40  
 deixar a cidade,6,26  
 deixar o ministério,7,24  
 deixar o poder,8,22  
 deixar o partido,9,20  
 deixar dúvidas,10,18  
 deixar o time,11,17  
 deixar a equipe,12,16  
 deixar o clube,13,16  
 deixar a fazenda,14,15 (COE)  
 deixar o campo,15,15  
 deixar a prisão,16,13  
 deixar vítimas,17,12  
 deixar o carro,18,12  
 deixar a favela,19,12  
 deixar a prefeitura,20,11  
 deixar suas casas,21,9 (COE)  
 deixar o futebol,22,9  
 deixar a casa,23,9 (COE)  
 deixar a empresa,24,9  
 deixar o hotel,25,9  
 deixar a seleção,26,9  
 deixar sua casa,27,8 (COE)  
 deixar filha,28,8  
 deixar o hospital,29,8  
 deixar marcas,30,8  
 deixar a sala,31,8  
 deixar a presidência,32,8  
 deixar o emprego,33,7  
 (deixar as coisas),34,7 (JAN)  
 deixar cargo,35,7  
 deixar sua marca,36,7(COE)  
 deixar a área,37,6  
 deixar seu país,38,6 (COE)  
 deixar o palmeiras,39,6  
 deixar o prédio,40,6 (COE)  
 deixar vestígios,41,6  
 deixar o apartamento,42,6  
 deixar a escola,43,5  
 deixar saudade,44,5  
 deixar o morro,45,5  
 deixar o gabinete,46,5  
 deixar as drogas,47,5  
 deixar seus cargos,48,5 (COE)  
 deixar dúvida,49,5  
 deixar o plenário,50,4  
 (deixar patente),51,4 (ETQ)  
 deixar a delegacia,52,4  
 deixar seqüelas,53,4  
 deixar o mercado,54,4 (COE)  
 deixar os cargos,55,4 (COE)  
 deixar o basquete,56,4  
 deixar a mulher,57,4  
 deixar barato,58,4  
 deixar as quadras,59,4  
 deixar o palco,60,4  
 deixar a quadra,61,4  
 deixar o aeroporto,62,4  
 deixar o gramado,63,4  
 deixar a universidade,64,3  
 deixar o estádio,65,3  
 deixar o fluminense,66,3  
 deixar o quarto,67,3  
 deixar essa posição,68,3 (COE)  
 deixar o vôlei,69,3  
 (deixar pistas),70,3 (JAN)  
 deixar o grupo,71,3 (COE)  
 deixar recados,72,3  
 deixar a emissora,73,3 (COE)  
 deixar o pai,74,3  
 deixar o jogo,75,3  
 deixar a política,76,3  
 deixar a clínica,77,3 (COE)  
 deixar seu voto,78,3 (COE)  
 deixar o senado,79,3 (COE)  
 deixar a embarcação,80,3  
 deixar a profissão,81,3  
 deixar a receita,82,3 (COE)  
 deixar Gaza,83,3  
 (deixar as pessoas),84,3 (JAN)  
 deixar rastros,85,3  
 deixar suas funções,86,3 (COE)

deixar os filhos,87,3  
(deixar abertas),88,3 (ETQ)  
deixar o estado,89,3  
deixar o palanque,90,3 (COE)  
deixar a propriedade,91,3 (COE)  
deixar a capital,92,3  
deixar a copa,93,3 (COE)  
(deixar os jogadores),94,3 (JAN)  
deixar o banco,95,3  
deixar a defesa,96,3  
deixar o presídio,97,3  
deixar testemunhas,98,3  
deixar o exército,99,3  
deixar um bilhete,100,3  
deixar o dinheiro,101,3 (COE)  
deixar o caso,102,2  
deixar os campos,103,2  
(deixar o resto),104,2 (JAN)  
deixar a legenda,105,2  
deixar o bairro,106,2

**Teste 8: TOMAR (DET) N**

- tomar posse,1,164  
 tomar banho,2,58  
 tomar decisões,3,52  
 tomar café,4,43  
 tomar uma decisão,5,33  
 tomar conhecimento,6,29  
 tomar providências,7,26  
 tomar cuidado,8,26  
 tomar o poder,9,22  
 tomar sol,10,22  
 tomar a decisão,11,20 (COE)  
 tomar a iniciativa,12,15 (COE)  
 tomar um banho,13,13  
 tomar gols,14,12  
 tomar conta,15,12  
 tomar a bola,16,12  
 (tomar medidas),17,11(JAN)  
 tomar chá,18,11  
 tomar cerveja,19,11  
 tomar partido,20,11  
 tomar tempo,21,10  
 tomar forma,22,9  
 tomar uma atitude,23,9  
 tomar drogas,24,9  
 (tomar posição),25,9 (JAN)  
 tomar alguns cuidados,26,9  
 tomar café=da=manhã,27,8  
 tomar remédios,28,8  
 tomar uma providência,29,8  
 (tomar as providências),30,8 (JAN)  
 tomar sorvete,31,8  
 tomar coragem,32,8  
 tomar a cidade,33,7 (COE)  
 tomar uma posição,34,7  
 tomar essa decisão,35,7 (COE)  
 tomar fôlego,36,6  
 tomar o remédio,37,6 (COE)  
 tomar água,38,6  
 tomar esta decisão,39,6 (COE)  
 tomar algumas precauções,40,6  
 tomar um táxi,41,6  
 tomar um ônibus,42,6  
 (tomar qualquer providência) 43, 6 (JAN)  
 tomar sua decisão,44,6 (COE)  
 tomar a mesma decisão,45,6 (COE)  
 tomar jeito,46,6  
 (tomar dinheiro),47,5 (JAN)  
 tomar um café,48,5  
 tomar vitaminas,49,5  
 tomar vinho,50,5  
 tomar a droga,51,5 (COE)  
 tomar corpo,52,5  
 (tomar parte),53,5 (JAN)  
 tomar um chope,54,5  
 tomar o gol,55,5 (COE)  
 tomar muito cuidado,56,5  
 tomar remédio,57,5  
 (tomar um rumo),58,4 (JAN)  
 tomar café-da-manhã,59,4  
 (tomar qualquer decisão),60,4 (JAN)  
 tomar um susto,61,4  
 tomar a palavra,62,4  
 tomar pílula,63,4  
 tomar gol,64,4  
 (tomar qualquer atitude),65,4 (JAN)  
 tomar a vacina,66,4 (COE)  
 tomar ônibus,67,4  
 tomar iniciativa,68,4  
 tomar champanhe,69,4  
 tomar álcool,70,4  
 tomar tal decisão,71,4 (COE)  
 tomar nota,72,4  
 tomar todas=as decisões,73,4  
 tomar nenhuma providência,74,3 (JAN)  
 tomar uma injeção,75,3  
 tomar mamadeira,76,3  
 tomar os medicamentos,77,3 (COE)  
 tomar um cafezinho,78,3  
 tomar alguma providência,79,3  
 tomar esta atitude,80,3 (COE)  
 tomar precauções,81,3  
 tomar providência,82,3  
 tomar chope,83,3

tomar um lanche,84,3  
tomar juízo,85,3  
tomar a dianteira,86,3  
tomar pílulas,87,3  
tomar notas,88,3  
tomar o seu lugar,89,3 (COE)  
tomar vergonha,90,3  
tomar muito=tempo,91,3  
tomar antidepressivos,92,3  
tomar aldeia,93,3 (JAN) (MAN)  
tomar comprimidos,94,3  
tomar sopa,95,3  
tomar um porre,96,3  
tomar a pílula,97,3 (COE)  
(tomar o azt),98,3 (DAT)  
tomar o microfone,99,3  
tomar veneno,100,3  
tomar empréstimo,101,3  
tomar leite,102,3  
tomar o café=da=manhã,103,3  
tomar o comprimido,104,2 (COE)  
tomar calmantes,105,2  
tomar essas providências,106,2 (COE)  
tomar um drinque,107,2  
tomar aspirina,108,2  
tomar alguma atitude,109,2  
tomar a frente,110,2  
tomar os cartões,111,2 (COE)

**Teste 9: GANHAR (DET) N**

ganhar dinheiro,1,101	ganhar mais força,42,5
ganhar a eleição,2,90	ganhar o campeonato,43,5
ganhar a copa,3,56	(ganhar versão),44,5 (JAN)
ganhar as eleições,4,42	ganhar o prêmio,45,5 (COE)
ganhar espaço,5,35	ganhar a presidência,46,5
ganhar tempo,6,34	ganhar alguma coisa,47,5
ganhar o jogo,7,32	ganhar medalhas,48,5
ganhar força,8,32	ganhar uma eleição,49,5
ganhar a vida,9,28	ganhar autonomia,50,5
ganhar terreno,10,16	ganhar presentes,51,5
ganhar jogo,11,16	ganhar a concorrência,52,5
ganhar muito dinheiro,12,15	(ganhar líquido),53,5 (ETQ)
ganhar votos,13,15	(ganhar financeiro),54,5 (ETQ)
ganhar importância,14,15	(ganhar contornos),55,4 (JAN)
ganhar destaque,15,12	ganhar adeptos,56,4
ganhar o título,16,12	ganhar o consumidor,57,4
ganhar eleição,17,12	ganhar o torneio,58,4
(ganhar real),18,11 (ETQ)	ganhar status,59,4
ganhar salário=mínimo,19,11	ganhar eleições,60,4
ganhar mercado,20,11	ganhar a posição,61,4
ganhar corpo,21,11	ganhar eficiência,62,4
ganhar peso,22,10	ganhar manchetes,63,4
ganhar fama,23,10	ganhar essa eleição,64,4 (COE)
ganhar títulos,24,9	ganhar o brasileiro,65,4(COE)
ganhar experiência,25,9	ganhar intensidade,66,4
ganhar a partida,26,9 (COE)	ganhar um título,67,4
ganhar um oscar,27,9	(ganhar linhas),68,4 (JAN)
ganhar prêmios,28,8	ganhar ritmo,69,4
ganhar uma copa,29,8	ganhar a licitação,70,4 (COE)
ganhar confiança,30,8	ganhar agilidade,71,4
ganhar impulso,31,8	ganhar vida,72,4
ganhar algum dinheiro,32,7	ganhar popularidade,73,4
ganhar o oscar,33,7	ganhar liberdade,74,3
ganhar mais dinheiro,34,7	ganhar a viagem,75,3
ganhar o mundo,35,7	ganhar a taça,76,3
ganhar a causa,36,6	ganhar vários títulos,77,3
ganhar a guerra,37,6 (COE)	ganhar volume,78,3
ganhar credibilidade,38,6	ganhar todas=as partidas,79,3
ganhar pontos,39,6	ganhar produtividade,80,3
ganhar fôlego,40,6	ganhar a liberdade,81,3
ganhar velocidade,41,6	ganhar salário,82,3

ganhar nome,83,3  
ganhar a mesma coisa,84,3 (COE)  
ganhar qualidade,85,3  
ganhar espaços,86,3  
ganhar corridas,87,3  
(ganhar motor),88,3 (JAN)  
ganhar salários,89,3  
ganhar troféus,90,3  
ganhar roupas,91,3  
ganhar respeito,92,3  
ganhar a parada,93,3  
ganhar a ação,94,3 (COE)  
ganhar um prêmio,95,3  
ganhar pouco dinheiro,96,3

ganhar a prova,97,3  
ganhar alguns quilos,98,3  
ganhar copa,99,3 (MAN)  
(ganhar seu nome),100,3 (JAN)  
ganhar o nobel,101,3  
ganhar prêmio,102,3  
ganhar animação,103,3  
ganhar forma,104,3  
(ganhar um número),105,2 (JAN)  
ganhar muita grana,106,2  
ganhar mais tempo,107,2  
(ganhar outra caravana),108,2 (JAN)  
ganhar uma partida,109,2  
ganhar a liderança,110,2

**Teste 10: VER (DET) N  
metatexto do corpus**

ver quadro,1,150 (MET)  
 ver texto,2,65 (MET)  
 ver o filme,3,39  
 ver o jogo,4,30  
 ver tv,5,29  
 ver computador,6,25  
 ver televisão,7,15  
 ver o crime,8,12  
 ver tabela,9,11 (MET)  
 ver a cena,10,11  
 ver a copa,11,11  
 ver as coisas,12,10  
 ver o mundo,13,10  
 ver o acidente,14,10  
 ver mapa,15,9 (MET)  
 ver navios,16,8 (JAN)  
 ver foto,17,7 (MET)  
 ver correio,18,7  
 ver modem,19,6  
 ver os policiais,20,6  
 ver o show,21,6  
 ver um filme,22,6  
 ver o espetáculo,23,6  
 ver as fotos,24,6 (MET)  
 ver fotos,25,5 (MET)  
 ver a moda,26,5  
 ver o quadro,27,5 (MET)  
 ver o lance,28,5  
 ver engenharia,29,5  
 ver a lista,30,5  
 ver desenho,31,5 (MET)  
 ver o carro,32,5  
 ver esse filme,33,5  
 ver coisas,34,5  
 ver o sol,35,4  
 ver um jogo,36,4  
 ver o rosto,37,4  
 ver a jardinagem,38,4  
 ver ilustração,39,4 (MET)  
 ver novelas,40,4  
 ver o filho,41,4  
 ver parque,42,4  
 ver estrelas,43,4  
 ver a vida,44,4  
 ver textos,45,4 (MET)  
 ver a cidade,46,4  
 ver o fenômeno,47,4  
 ver problemas,48,4  
 ver o resultado,49,4 (MET)  
 ver crediário,50,4  
 ver gente,51,4  
 ver o país,52,4  
 ver os ladrões,53,4  
 ver essas coisas,54,4  
 ver filmes,55,4  
 ver os corpos,56,4  
 ver o marido,57,4  
 ver reportagem,58,4 (MET)  
 ver futebol,59,3  
 ver a luz,60,3  
 ver as imagens,61,3  
 ver o cinema,62,3  
 ver o time,63,3  
 ver a coisa,64,3  
 ver o discurso,65,3  
 ver o trailer,66,3  
 ver o papa,67,3  
 ver sangue,68,3  
 ver o roubo,69,3  
 ver arte,70,3  
 ver alguma coisa,71,3  
 ver a página,72,3 (MET)  
 ver o gol,73,3  
 ver uma coisa,74,3  
 ver livro,75,3  
 ver endereços,76,3  
 ver os preços,77,3  
 ver os problemas,78,3  
 ver as pessoas,79,3  
 ver uma cena,80,3  
 ver a seleção,81,3

ver o coração,82,3  
ver seus filmes,83,3  
ver preços,84,3 (MET)  
ver a tv,85,3  
ver os dados,86,3 (MET)  
ver a uva,87,3 (JAN)  
ver o evento,88,3  
ver importação,89,3  
ver pessoas,90,3  
ver o problema,91,3  
ver relação,92,3 (MET) (JAN)  
ver tanta gente,93,3  
ver um homem,94,3  
ver a noite,95,3  
ver dificuldades,96,3  
ver a mulher,97,3  
ver a televisão,98,2  
ver o contrato,99,2  
ver pó,100,2

**Teste 11: CRIAR (DET) N**

criar a urv,1,173 (DAT)  
 criar o real,2,64  
 criar empregos,3,50  
 criar problemas,4,18  
 criar jogadas,5,9  
 criar polêmica,6,8  
 (criar condições),7,8 (JAN)  
 criar mais empregos,8,7  
 criar um clima,9,7  
 criar chances,10,6  
 criar um estilo,11,6  
 criar distorções,12,5  
 criar o indexador,13,5  
 (criar um fato),14,5(JAN)  
 criar jurisprudência,15,5  
 criar produtos,16,5  
 criar várias chances,17,4  
 criar o plano,18,4 (COE)  
 (criar ambientes),19,4 (JAN)  
 (criar este ano),20,4 (JAN)  
 criar cenários,21,4  
 criar meus filhos,22,4  
 criar histórias,23,4  
 criar obstáculos,24,4  
 criar gado,25,4  
 criar dependência,26,4  
 criar coragem,27,4  
 criar uma linguagem,28,4  
 criar inflação,29,4  
 criar personagens,30,4  
 criar asas,31,4  
 criar situações,32,4  
 criar o fundo,33,3 (COE)  
 criar uma expectativa,34,3  
 criar o psdb,35,3  
 criar confusões,36,3  
 criar a peça,37,3 (COE)  
 criar fama,38,3  
 criar moeda,39,3 (MAN)  
 criar desenhos,40,3  
 criar a moeda,41,3 (COE)  
 criar regras,42,3  
 criar outro problema,43,3 (COE)

criar os filhos,44,3  
 criar atritos,45,3  
 (criar um mundo),46,3 (JAN)  
 (criar mensagens),47,3 (JAN)  
 criar leis,48,3  
 criar espaços,49,3  
 (criar uma relação),50,3 (JAN)  
 (criar santos),51,3 (JAN)  
 criar expectativas,52,3  
 (criar obras),53,3 (JAN)  
 (criar tensões),54,3 (JAN)  
 criar postos=de=trabalho,55,3  
 criar um personagem,56,3  
 criar alternativas,57,3  
 criar ilusões,58,3  
 (criar um espaço),59,3 (JAN)  
 criar o serviço,60,2 (COE)  
 criar movimentos,61,2  
 criar desemprego,62,2  
 criar oportunidades,63,2  
 criar sindicatos,64,2  
 criar as jogadas,65,2  
 criar raízes,66,2  
 (criar a mulher),67,2(JAN)  
 criar um nosso estilo,68,2 (COE)  
 criar uma imagem,69,2  
 criar o universo,70,2  
 criar outros fundos,71,2  
 criar um gênero,72,2  
 criar drinques,73,2  
 criar outras chances,74,2  
 criar imagens,75,2  
 criar ovelhas,76,2  
 criar impostos,77,2  
 criar um boletim,78,2  
 (criar um sujeito),79,2 (JAN)  
 criar uma jurisprudência,80,2  
 (criar valores),81,2 (JAN)  
 criar a x-girl,82,2  
 criar o futuro,83,2  
 criar os anúncios,84,2 (COE)  
 criar dificuldades,85,2  
 (criar receita),86,2 (JAN)  
 criar punições,87,2

criar um ambiente,93,2  
criar a coreografia,101,2 (COE)  
criar a cena,102,2 (COE)  
criar roupas,103,2  
criar pânico,104,2  
criar tais juizados,105,2 (COE)  
criar confusão,106,2  
(criar uma sociedade),107,2 (JAN)  
criar o clima,108,2 (COE)  
criar um desequilíbrio,109,2  
(criar colônias),110,2 (JAN)  
criar uma cultura,111,2  
criar despesas,88,2  
criar indisposições,89,2  
criar pôsteres,90,2  
criar galinhas,91,2  
criar cargos,92,2

criar um cenário,94,2  
criar museus,95,2  
(criar mecanismos),96,2 (JAN)  
criar tumulto,97,2  
criar direitos,98,2  
criar ruído,99,2  
criar algumas jogadas,100,2  
criar uma forma,112,2  
criar palavras,113,2  
criar uma apresentação,114,2  
(criar alguma peça),115,2 (JAN)  
criar um debate,116,2  
criar o problema,117,2 (COE)  
criar tributos,118,2  
criar animais,119,2  
criar os modelos,120,2 (COE)  
criar ansiedade,121,2