

Padrões da Preposição “de” entre Sintagmas Nominais em Linguagem Cotidiana e Linguagens Técnico-Científicas¹

Aline Villavicencio ¹, Maria José B. Finatto ², Viviane Possamai²

¹ Instituto de Informática da UFRGS,
Porto Alegre –RS, Brasil
villavicencio@inf.ufrgs.br

² Instituto de Letras da UFRGS,
Porto Alegre - RS, Brasil
mfinatto@terra.com.br
vivianepossamai@yahoo.com.br

Abstract. The preposition DE is the most frequent written word in Brazilian Portuguese. However, in spite of this, there seems to be a lack of investigation on its occurrence behavior and distributional properties. In this paper we attempt to move one step forward in the profiling of this word, by investigating its description in some grammatical handbooks, dictionaries and its occurrence in corpora. We observe that DE collocates much more frequently with names than with verbs in technical or specialized corpora. At the end, we focus on its use preceded by substantives.

1 Introdução

A preposição DE é a palavra com maior frequência de uso no português escrito do Brasil. Em função de sua multiplicidade e quantidade de empregos, observá-la em corpora com o fim de depreender suas padronizações de uso e distribuição exige que se esteja preparado para lidar com grandes quantidades de dados. Sua alta frequência, mesmo que examinada apenas em amostragens de corpora, ainda tende a representar um desafio importante, quer pela quantidade de dados, quer pela obtenção de um adequado balanceamento ou desenho de um recorte representativo para o estudo de seus usos na língua.

Paradoxalmente, ao mesmo tempo em que há tão alta frequência de uso da partícula, parece que são raros os estudos específicos sobre seus padrões ou sobre sua alta incidência e larga distribuição na língua, uma característica que se revela saliente em qualquer acervo textual que se observe. Este trabalho pretende, modestamente, contribuir para que tenhamos à disposição alguns dados, ainda que de caráter exploratório, sobre padrões de funcionamento de DE no português brasileiro. O

¹ Este estudo aproveita dados de pesquisa realizada com apoio do CNPq e da FAPERGS ao projeto TEXTQUIM. A participação no evento V Encontro de Corpora só foi possível graças ao auxílio do Programa de Pós-Graduação em Letras e Pró-Reitoria de Pesquisa da UFRGS.

objetivo é que possamos ter alguma base para avançar o patamar da informação, geralmente sucinta, que nos é fornecida por gramáticas tradicionais e dicionários gerais da língua portuguesa.

A partícula DE, por sua natureza e pela feição gramatical de uma língua românica como é o português, funciona como um elemento associativo. Assim, podemos imaginar que opere tal como uma “ponte” ou “cola” que liga diferentes tipos de palavras entre si: substantivos, verbos, adjetivos, advérbios, etc. Essa “ponte” é um elemento imprescindível para o funcionamento sintático e semântico da língua, de modo que um padrão preposicional mostra um traço intrínseco ao padrão gramatical do português e pode inclusive caracterizar essa língua frente a outras.

Vejamos um exemplo de língua em que a preposição não é o elemento mais freqüente: em uma língua germânica como inglês, a palavra mais freqüente é um determinante, a palavra THE. O contraste entre DE e THE como palavras mais freqüentes em cada caso já nos fornece um pequeno retrato instantâneo de indicativos sobre as diferenças mais básicas entre essas duas línguas. O português, nesse retrato, ainda que um tanto desfocado, aparece, à primeira vista, como uma língua de associações enquanto o inglês mostra-se como uma língua de determinações.

A título de uma primeira aproximação sobre a função “vinculadora” de DE, buscamos um panorama geral sobre sua incidência em diferentes corpora. A partir de um reconhecimento genérico de suas combinatórias mais usuais, destacamos, ao final do trabalho, sua associação a nomes, especificamente o caso da preposição precedida por substantivos, o que corresponde ao padrão NOME+DE+OUTROS ELEMENTOS NOMINAIS em uma cadeia sintagmática. Nosso foco maior de interesse incidirá, entretanto, apenas sobre o segmento inicial NOME+DE dessa cadeia.

Além de relatar nossos achados gerais sobre a freqüência de DE em corpora de tipo geral e em corpora de algumas linguagens técnico-científicas, trazemos aqui também informações coletados em fontes de referência: dicionários da língua portuguesa e manuais de gramática. Depois de percorrer dicionários e gramáticas, observamos seu tratamento no léxico do português brasileiro feito pelo NILC². Ao final do estudo, tecemos algumas considerações sobre os padrões observados para DE entre sintagmas nominais e discutimos o aproveitamento dos dados obtidos.

2 DE em corpora: visão geral

A alta freqüência dessa preposição é confirmada em diversos corpora de linguagem escrita, seja de registros de linguagem cotidiana ou de linguagem técnica/especializada. Um primeiro exemplo de confirmação vemos no Banco de Português³ (BP), um corpus de tipo geral atualizado constantemente. Entre 2004 e 2005 atingiu 240 milhões de palavras.

² Núcleo Interinstitucional de Lingüística Computacional da Universidade de São Paulo, com sede em São Carlos - SP.

³ <http://www2.lael.pucsp.br/corpora/bp/index.htm>

Em 2003, esse acervo continha 223 milhões de palavras. Foi criado e é mantido no âmbito do projeto DIRECT, junto ao LAEL da PUC-SP. As 10 palavras mais freqüentes desse corpus, considerando dados de 2003, estão na Tabela 1 a seguir:

Tabela 1. As 10 palavras mais freqüentes no Banco de Português (BP) – ano de 2003.

ranking	palavra	ocorrências	percentual	ranking	palavra	ocorrências	percentual
1	DE	1.537.460	4,42	6	DO	609.521	1,75
2	A	1.082.233	3,11	7	DA	545.271	1,57
3	O	1.026.380	2,95	8	EM	443.567	1,28
4	E	726.548	2,09	9	PARA	353.847	1,02
5	QUE	667.850	1,92	10	NO	308.932	0,89

DE ocupa a mesma primeira posição em vários outros corpora, tal como o Corpus Folha⁴, com 223 milhões de palavras, Tabela 2 (a), e o corpus NILC⁵, com 40 milhões de palavras, Tabela 2(b). Veja-se que o percentual de DE em cada um desses corpus é bastante aproximado, 4,7 e 4,2 % respectivamente.

Tabela 2. As cinco palavras mais freqüentes no Corpus Folha e no Corpus NILC (segmento corpus corrigido)

Corpus Folha (a)		Corpus NILC (b)	
ranking	palavra	ranking	palavra
1	DE (4,7%)	1	DE (4,2%)
2	A	2	A
3	O	3	O
4	E	4	E
5	QUE	5	QUE

Observa-se padrão semelhante de freqüências para essas cinco palavras também em corpora especializados, compostos por textos técnicos e científicos de diferentes gêneros.

Um exemplo de corpus especializado é Corpus TEXTQUIM⁶. Esse corpus tem 1 milhão de palavras e é integrado apenas por textos de Química: textos didáticos do ensino superior, textos científicos *stricto sensu* – artigos de periódico especializado - e textos de popularização de temas de ciência.

⁴ O CETENFolha (Corpus de Extractos de Textos Eletrônicos NILC/Folha de S. Paulo) é um corpus de cerca de 24 milhões de palavras em português brasileiro com base nos textos do jornal Folha de São Paulo que fazem parte do corpus NILC/São Carlos, compilado pelo NILC. Disponível em <http://acdc.linguatca.pt/cetenfolha>.

⁵ Corpus NILC, parte corrigida. Disponível em <http://www.nilc.icmsc.sc.usp.br/nilc/tools/corpora.htm>

⁶ Dados disponíveis em www.ufrgs.br/textquim

Em um recorte do corpus TEXTQUIM, com manuais didáticos acadêmicos, 368.005 palavras, verificamos 18.864 ocorrências de DE (5,12%) e 14.302 ocorrências de A (3,88%), palavra de segunda maior ocorrência.

Em outro segmento desse mesmo corpus, composto apenas por artigos da Revista Química Nova, uma revista científica especializada, temos, em uma amostra de 127.871 palavras, 6.980 ocorrências de DE (5,45%).

Num outro corpus especializado, o CORTEC⁷, no segmento dedicado ao tema ECOTURISMO, composto por textos originais retirados de sites do governo brasileiro, de entidades ambientalistas e de agências de turismo, temos uma amostra de textos com 200.887 palavras. Nessa amostra, há 11.850 ocorrências de DE, o que representa 5,8%. Vale notar que há apenas um verbo entre os 100 primeiros contextos de DE gerados por ferramenta do próprio site: entidade que acaba DE nascer para.

No que tange às frequências e associações no vasto ambiente da WEB visto como um grande corpus, foram observados usos de DE em *sites* brasileiros com a ferramenta Webcorp. Em uma busca realizada em 02/01/2006, tomamos como referência a ferramenta Google.

Como a presença de DE é muito grande, depois de quase uma hora de varredura, obtivemos apenas um resultado parcial. Ao acessar 200 páginas web, 16 com erro, foram gerados pelo Webcorp 3.110 contextos. Nesses contextos, o padrão majoritário de distribuição da preposição é o de ser antecedido por expressões nominais. Nesse pequeno universo de 3.110 usos, a primeira ocorrência de VERBO+DE é encontrada apenas depois dos 150 primeiros contextos.

Os primeiros usos de DE nessa busca, foram os seguintes:

1. Universidade Federal de Santa Catarina;
2. Departamento de Engenharia;
3. bolsa de produtividade em pesquisa;
4. comprovantes de rendimento;
5. Rio de Janeiro;
6. páginas de links;
7. proposta de implantação;
8. campus de Limeira.

Em outra situação de buscas, no corpus Lácio Web⁸, um corpus cuja consulta é oferecida por tipos de etiquetagem, compusemos uma amostra da parte denominada Macmorph (corpus anotado morfossintaticamente). Nossa amostra compreendeu 1.167.183 palavras e foi composta por textos jornalísticos de dez cadernos do jornal

⁷ O CorTec - Corpus Técnico-Científico - é um corpus comparável de textos técnicos e/ou científicos originalmente escritos em português brasileiro e em inglês. Inclui as seguintes áreas: Culinária, Hipertensão, Informática e Instrumentos Contratuais. Disponível em <http://www.fflch.usp.br/dlm/comet>.

⁸ O Lácio Web é um site brasileiro com um corpus contemporâneo de língua geral, subdividido em subcorpora que representam vários gêneros e tipos textuais - textos disponíveis para download e para serem usados com as próprias ferramentas do site. <http://www.nilc.icmc.usp.br/lacioweb/index.htm>

Folha de São Paulo, ano 1994. Desse recorte retiramos um conjunto que corresponde a todas as 82.372 ocorrências de DE identificadas pela etiqueta PREP (preposição).

Considerando-se, por amostragem, as 100 primeiras ocorrências de DE PREP nesse corpus etiquetado, repete-se o padrão majoritário de NOMES+DE. Há verbos antecedendo DE apenas em dois usos entre os 100 primeiros listados: cobrar DE+o ministro, o alerta é DE+o pesquisador. Nesses 100 contextos, temos, além desses poucos verbos, 07 adjetivos antecedendo DE e 02 formas participiais. A maioria absoluta é de substantivos na posição que antecede a preposição.

No mesmo corpus, agora no seu segmento Lácio-Ref, com 7.935.618 palavras, um corpus de tipo geral não anotado, verificamos que DE também é a palavra de maior ocorrência, com 362.841 registros, o que nos dá um percentual relativo de 4,5%. A palavra A ocupa a segunda posição com 293.499 ocorrências (3,6%).

Nesses corpora, Banco de Português, Corpus Folha, Corpus NILC, TEXTQUIM, CORTEC-Ecoturismo e Lácio-Ref, o percentual levemente variável de DE, embora permaneça fixo na primeira posição do ranking das palavras mais frequentes, delinea uma questão “bruta” que pode ser investigada: quanto mais técnico ou específico fosse um corpus, maior seria o número de DE? Trataremos de ajustar ou refinar a formulação dessa questão mais adiante.

Independentemente da feição da pergunta, o que se confirma, sem dúvida, nessas observações, é que DE é a palavra mais repetida com percentual de ocorrência entre 4,5 e 5,8% em relação ao número total de palavras dos corpora examinados. Além disso, é uma partícula muito mais antecedida por nomes do que por verbos.

3 DE em corpora: linguagem comum e especializada

Buscando depreender algum padrão de diferença para DE em linguagem comum e linguagens “técnicas”, fizemos novos contrastes entre corpora.

Examinamos, primeiro, o subconjunto do Banco de Português (BP) disponível para consulta *on-line*, com 1,1 milhão de palavras. Desse total, foram gerados 16 mil contextos pelo listador de concordâncias oferecido no próprio site. A partir desses 16 mil contextos, examinamos os 2.603 *clusters*⁹ com DE com repetição igual ou superior a seis vezes. É importante salientar que essa amostra do BP de 1,1 milhão de palavras corresponde a um corpus de tipo geral, bem balanceado por tipos, composto por vários gêneros textuais. Esse segmento consultado inclui textos especializados como manuais, teses, dissertações e cartas comerciais, além de textos literários e textos jornalísticos.

Na nossa lista de *clusters*, repete-se a presença majoritária de DE antecedido por nomes. DE teve antecedente verbal em apenas 7,3% dos 2.603 casos, tais como lembrar de, fugir de e voar de. A vasta maioria dos casos de *clusters* com DE corresponde ao padrão NOME+DE, tal como em gerente de produção e taxa de transferência.

⁹ Clusters são aglomerados reiterados de palavras em torno de uma palavra que se observe. Esses clusters foram obtidos pela aplicação do software Wordsmith tools às 16.000 ocorrências geradas pelo listador de contextos disponível no site do BP.

Pensando na necessidade de um maior detalhamento por tipos de textos das freqüências de DE, desenhamos um pequeno panorama de freqüências a partir de diferentes amostras de alguns corpora. A Tabela 3 a seguir mostra ocorrências de DE nessas amostras que têm diferentes tipos de texto, mais e menos especializados, bem como o percentual de ocorrência.

Tabela 3. Freqüências de DE em diferentes amostra de corpora

corpus/segmento/tipos de texto	dimensão da amostra	ocorrências	Percentual em ordem decrescente de ocorrências por 100 palavras.
CORTEC/ Culinária/ receitas	252.149	22.093	8.7
TEXTQUIM/ artigos da revista Química Nova/ Amostra 1	415.847	27.277	6.5
TEXTQUIM/Manual de Físico-Química (capítulos sobre Termodinâmica) [5]	54.625	3.013	5.5
TEXTQUIM/artigos da revista Química Nova Amostra 2	128.729	7.028	5.4
BP/textos de jornais	199.285	9.607	4.8
CDs Folha de SP/jornais	223.371.280	10.491.746	4.6
BP/ geral-misto vários tipos de texto	1.182.993	52.326	4.4
CORTEC/ Informática/textos jornalísticos – revista Infoexame e caderno de Informática da Folha de SP	207.358	8.434	4.0
POSSAMAI/ artigos de comunicações em eventos da SBC [6]	1.287.260	16.000	1.2

Pelo que se pode observar, no ranking de DE separado em grandes tipos de corpora, vemos que os textos da área de Informática são os que apresentam menos ocorrências de DE, o que sugere uma investigação mais a fundo dos motivos para tamanha disparidade com os textos que se encontram no primeiro lugar, por exemplo. Um dos

motivos mais evidentes é que a Culinária e suas expressões “colher de”, “copo de” faz crescer exponencialmente o número de ocorrências da partícula.

Caracterizando um pouco mais os outros corpora presentes nessa tabela, temos, além dos já referidos CORTEC (Culinária) e Possamai, no tema “Informática”, com artigos que correspondem a comunicações apresentadas em eventos da Sociedade Brasileira de Computação (SBC).

Os três corpora representando amostras da área de Química ficaram com percentuais muito semelhantes na análise. Um recorte composto apenas por artigos do periódico Química Nova, uma segunda amostra, maior que a primeira, e uma terceira composta apenas pelo texto de um manual acadêmico de Físico-Química nos capítulos que tratam sobre Termodinâmica.

Em posição intermediária na tabela, vemos um segmento do corpus BP formado apenas por textos de jornais, o corpus de CDs da Folha de São Paulo organizado por Berber Sardinha [1] e o BP geral.

Em penúltimo, novamente o corpus CORTEC, no segmento “Informática”, salientando-se que, diferente do corpus Possamai, é composto por textos jornalísticos de popularização de Informática, coletados da revista Infoexame e do Caderno de Informática do jornal Folha de São Paulo.

O conjunto de textos e os dados levantados ainda não permitem fazer uma avaliação da incidência do uso de DE conforme o caráter mais ou menos especializado do texto. Para tal conclusão, mais interessante seria observar os contextos de colocação e co-ocorrência de DE para identificar por que foram tão freqüentes em culinária e tão menos freqüentes em Informática. O que está posto, porém, é que os padrões de constituição lexical das diversas áreas de conhecimento apresentam diferenças. O estudo poderia prosseguir avaliando que estruturas a Culinária utiliza que tanto fazem crescer a incidência de DE e, como contraponto, que estruturas a Informática acadêmica utiliza diferentes das de Culinária, ou deixa de utilizar.

4 DE em gramáticas e dicionários: uma realidade diferente

4.1 Gramáticas

Trazemos agora, como exemplo de nossas observações, uma breve descrição das informações sobre DE disponíveis em duas gramáticas. A primeira delas pode ser considerada um exemplo de obra tradicional, de preocupação prescritiva. A segunda gramática tem um caráter bem mais moderno, é descritiva e baseada em corpora. A primeira obra é a *Gramática Normativa da Língua Portuguesa* de Rocha Lima [7] (ROCHA LIMA, 21ª edição, 1980). A outra é a *Gramática de Usos do Português* de Moura Neves [4], uma das poucas gramáticas produzidas no Brasil a partir da observação da língua em um corpus.

Segundo Rocha Lima, a preposição DE é uma preposição “fraca”, pois seria vazia de sentido. As suas funções, conforme o autor, são as de:

1. Introduzir complemento relativo de muitos verbos, tal como ocorre em precisar DE;
2. Iniciar objeto direto preposicional, como no exemplo Ouvirás DOS contos;
3. Preceder oração subordinada substantiva; sendo um exemplo citado pelo autor Dói-me DE ser eu;
4. Expressar relações: lugar onde, origem, causa, efeito: tal como no exemplo vir DE casa;
5. Ligar um substantivo a outro. Aqui reaparece a indicação de que DE expressa as relações antes mencionadas, tal como em mesa DE madeira.

É interessante observar que a primeira informação, nesta obra de Rocha Lima, é sobre o uso de DE antecedido de verbos do tipo transitivo indireto. Além disso, vemos um reconhecimento da capacidade expressiva de relações de sentido via DE, o que parece chocar-se com as colocações iniciais sobre a condição da partícula ser “fraca” ou vazia de sentido. Afinal, indica que DE expressa relações.

Por outro lado, na obra de Moura Neves, há muito mais informações sobre DE. Vemos, em síntese, dois grandes blocos de funções atribuídas para a partícula (os grifos são nossos):

1. DE funciona no sistema de transitividade, isto é, introduz complemento. Suas funções são a de introduzir complemento de:verbo;

- adjetivo: Afastado DE casa; livre DE pena;
- substantivo: O ataque DE holandeses;
- advérbio: Depois DO almoço.

2. DE funciona fora do sistema de transitividade, estabelecendo relações semânticas. Estabelece os seguintes tipos de relações:

- no sintagma verbal (adjunto adverbial): verbo+DE+sintagma nominal; exemplo: Ele agradece e vai saindo DE fininho.
- no sintagma nominal (adjunto adnominal): nome avalente+DE+sintagma nominal. Roubaram a bolsa DE uma mulher.

Além dessas relações, DE

- introduz sintagma em função predicativa (predicativo): O teto é DE zinco.
- integra construções indicativas de circunstâncias: Agia DE acordo com ela
- entra na construção de perífrases (+infinitivo): O doutor não cessa DE gracejar.
- constrói-se com o verbo ser+adjetivo disfórico+ DE+infinitivo. Exemplo: Vê se eu sou besta DE sustentar homem.
- entra em expressões fixas como dar DE ombros.

Conforme se percebe, as duas obras, em que pese o maior detalhamento da segunda, trazem, em primeiro lugar o funcionamento de DE como um elemento sintático, introdutor de um objeto indireto.

Naturalmente, Moura Neves distingue-se de Rocha Lima por fazer uma diferença de operação da partícula em dois níveis, dentro e fora do sistema de transitividade, o que

marca uma filiação teórica específica. No plano de “fora da transitividade”, isto é, de “fora” da sintaxe, vemos um território que, pressupomos, seja o da semântica. Nesse plano, DE, além de vinculador sintático, é um vinculador semântico, de modo que associa a significação estabelecida entre um antecedente e seu conseqüente. Não há, entretanto, qualquer menção sobre a alta freqüência de uso da preposição, tampouco há destaque para a maior freqüência de usos de DE fora da transitividade.

De qualquer modo, conforme vimos nos corpora, o padrão mais recorrente de DE, que é o de ser antecedido por nomes, não é o primeiro tipo de informação sobre o uso da partícula em nenhuma das duas gramáticas. Ambas parecem presas ao registro, em primeiro lugar, de uma função que parece ser a mais prototípica para DE. Essa função prototípica ou canônica é a de introdutor de complemento de verbo, ainda que não seja tão freqüente. Em uma gramática de usos, essa seria uma informação importante a registrar.

4.2 Dicionários tradicionais

No dicionário Houaiss, versão CD-ROM [3], e no Dicionário Aurélio Eletrônico Século XXI [2], que correspondem aos dicionários, em versão impressa, mais conhecidos da língua portuguesa, temos, em síntese, as seguintes informações sobre DE:

HOUAISS:

1. Relaciona palavras por subordinação: vive DE rendas. 2. Liga dois substantivos diretamente ou com auxílio de verbos 3. Introduz Objeto Direto preposicionado.

AURÉLIO XXI:

Partícula de larguíssimo emprego (grifo nosso). Usa-se: 1. Entre dois substantivos 2. Introduz o complemento terminativo de alguns verbos, adjetivos e substantivos (...) 3. Com auxiliares ter e haver e o infinitivo impessoal de outros, forma locuções perifrásticas do futuro: Hei de vencer (...).

Nesses dois dicionários a apresentação dos usos de DE está em ordem inversa à das gramáticas antes referidas. Primeiro vemos DE ligando nomes. Só depois é apresentada uma informação sobre o padrão de DE antecedido de verbo transitivo indireto. Essa ordem é mais compatível com a freqüência que verificamos em corpora.

Cabe sublinhar que dicionário Aurélio XXI é o único a trazer alguma indicação, entre os diversos materiais que consultamos, sobre a alta freqüência de uso na língua. Devemos ressaltar, também, que o uso de corpora e ferramentas informatizadas na produção de dicionários e gramáticas é ainda pouco difundido no Brasil, a se comparar com países como a Inglaterra, por exemplo. Os dados que apresentamos acima demonstram que a utilização de corpora pode vir a contribuir significativamente para a produção de materiais de referência que, de fato, representem a língua em uso e seus padrões, informações de alta relevância para seus usuários.

4.3 Dicionários computacionais: o léxico do NILC

O léxico do NILC¹⁰ contém 1.528.590 entradas. Dessas, 49.551 são nomes e 1.418.174 são verbos. No conjunto dos nomes, 1.762 registros aparecem com possibilidade de vinculação a DE, um exemplo é o substantivo abandono, que pode se combinar com DE, A ou com EM. Há 208 entradas de nomes que ocorrem exclusivamente com DE.

No conjunto dos verbos, que é bem maior do que o dos substantivos, temos apenas 363 verbos que ocorrem exclusivamente com DE e 1.194 verbos associados com DE de modo opcional, tal como abastecer DE/COM.

A proporção entre substantivos e verbos que se combinam exclusivamente com DE e o número de formas registradas no léxico atesta que DE se combina muito mais com nomes do que com verbos. Entretanto, no universo dos substantivos, são bem poucos os substantivos no padrão NOME+somente DE. O mais usual é que os nomes que antecedem DE também se combinem com outras preposições.

5 Padrões de DE entre sintagmas nominais

Para tentar depreender algum padrão na distribuição de DE entre sintagmas nominais, fizemos uma observação centrada nos clusters de DE mais freqüentes em corpora de linguagem comum e especializadas.

Em uma listagem com os mil primeiros clusters do BP, ordenados por freqüência, que inclui repetições que vão de 174 a 09 vezes do grupo de palavras, verificamos que expressões como: estado de São Paulo (174 repetições), formação de gerentes (77), transferência de calor (81), administração de empresas (58), gerente(s) de loja (122) e gerente de produto (112) estão entre as mais repetidas.

No corpus Possamai, nos 600 primeiros clusters com DE, temos como construções mais freqüentes, entre 81 e 50 repetições, largura de banda (81 repetições), banco de dados (80), sistemas de informação (64), Rio de Janeiro (61), base de dados (56), base de casos (55), qualidade de voz (52) e ponto de vista (50).

No corpus CORTEC-Culinária, que tem a maior presença de DE frente a diversos que examinamos, os clusters mais freqüentes são farinha de trigo (803 repetições), creme de leite (673), azeite de oliva (569), suco de limão (475), colheres de sopa (375) e modo de preparo (314). Entre os clusters de menor freqüência, na faixa entre 50 e 10 repetições, temos essência de baunilha (53), folha de louro (47), chá de molho (34), chá de mel (32), pão de forma (28), xícara de farinha (27), cubinhos de caldo (16), kg de açúcar (15), cheia de farinha (14), fatia de pão (12) e batedor de arame (10).

Pelo que é possível perceber, em Culinária o patamar de repetições de clusters no padrão nome+DE+nome é completamente distinto dos demais. No BP o diferencial em relação aos demais é que há mais nomes deverbais antecedendo DE.

Assim, pelo que se pode concluir, o padrão predominante de DE entre sintagmas nominais, em um corpus de tipo geral, é o de conector e o de introdutor de um nome

¹⁰ Agradecemos à gentil colaboração da Profa. Maria da Graça Volpe Nunes, que nos cedeu uma cópia do material para observação neste estudo.

que completa outro nome, semelhante ao papel de introdutor de complemento de verbo, pois temos nomes deverbais.

Nas linguagens especializadas, de Culinária e de Informática, pelo que podemos perceber, o padrão de DE é o de associar um nome e seu predicativo (por exemplo, farinha de trigo= X que é feito de Y) e há baixa incidência de nomes deverbais que antecedem a preposição.

6 Conclusões e considerações finais

Confirmamos que a preposição DE é a palavra escrita com maior uso no português do Brasil, sendo muito mais antecedida por nomes do que por verbos.

O padrão mais recorrente de DE, independente do tipo de corpora, é o NOME+DE+NOME. Pelos dados observados, pode-se concluir que o padrão predominante de DE entre sintagmas nominais, em um corpus de tipo geral, é o de conector/predicador (como no sintagma estado de São Paulo) e o de introdutor de um nome que completa outro nome (como no sintagma formação de gerentes), em uma função semelhante ao papel de introdutor de complemento de verbo, pois temos nomes deverbais.

Nas linguagens especializadas, o padrão de DE revela um item que associa um nome e seu predicativo (por exemplo, farinha de rosca) e uma baixa incidência de nomes deverbais precedendo a preposição.

No que se refere às gramáticas e dicionários da língua, vimos que há nas gramáticas um descompasso maior entre o que é mais utilizado em corpora e o que é registrado em primeiro lugar como função da preposição. Ao que parece, a função canônica de DE, um introdutor de complemento verbal, nos moldes de precisar de, é uma construção de baixa frequência.

Além das características já indicadas, acreditamos que é importante continuar estas observações para obter um quadro do funcionamento de DE também em sintagmas terminológicos (em termos químicos compostos por mais de uma palavra como, por exemplo, copo de Becker ou constante de equilíbrio), em expressões cristalizadas ou em *multiword expressions* como ponto de vista, tanto em linguagem especializada quanto em linguagem cotidiana. Vale [8], em um estudo sobre expressões cristalizadas do português integradas por verbos, verificou que a segunda maior frequência de preposições nessas construções é a da preposição DE. Temos aqui já um bom indicativo para continuidade desta exploração.

Outro aspecto que chamou nossa atenção, em meio às observações de verbos que antecederam DE no léxico do NILC, e que poderia ser melhor examinado, é que várias vezes o DE que está depois de um verbo não antecede um complemento verbal. O que verificamos foi uma inversão, algo semelhante à seqüência cortou DE João o benefício.

Independentemente das diferentes opções de prosseguimento deste nosso primeiro trabalho sobre o tema, acreditamos que tanto a alta frequência da preposição quanto o descompasso aparente entre um tipo de uso mais elevado e as informações de obras de referência já sinalizam a necessidade de continuarmos a investigar o tema. Afinal, a verificação de uma simples inversão de posição de DE entre as palavras mais

freqüentes de um texto já poderia indicar alguma provável “anomalia”. Se imaginarmos uma situação em que DE seja, não a primeira, mas a terceira palavra mais repetida num texto, tal que fosse um texto com muitas ocorrências de A e O enquanto o normal seria que fosse de DE, isso sinalizaria algum problema de construção. Enfim, se DE não for a palavra mais repetida, muito provavelmente haverá “algo estranho” ocorrendo na estruturação gramatical. Isso, entretanto, só pode ser percebido através de uma observação extensiva, feita em corpora e com apoio informatizado.

Referências

1. Berber Sardinha, Tony. A Língua Portuguesa nos CD-ROMS da Folha de São Paulo. *In*: BERBER SARDINHA, Tony. A Língua portuguesa no Computador. São Paulo: Mercado de Letras, 2005. p.215-248.
2. Ferreira, Aurélio Baurque de Holanda. DICIONÁRIO AURÉLIO ELETRÔNICO. NOVO DICIONÁRIO AURÉLIO - SÉCULO XXI. São Paulo: Nova Fronteira/Lexikon Informática, 1999.
3. Houaiss, Instituto Antônio. Dicionário Eletrônico Houaiss da língua portuguesa. Rio de Janeiro: Editora Objetiva, 2001.
4. Moura Neves, Maria Helena. Gramática de usos do português. São Paulo: UNESP, 2 edição, 2003.
5. Pilla, L. Físico-Química. Vol. 1. Rio de Janeiro: Livros Técnicos e Científicos, 1979.
6. Possamai, Viviane. Marcadores textuais do artigo científico em comparação português-inglês – um estudo sob a perspectiva da tradução. Porto Alegre: Universidade Federal do Rio Grande do Sul/Programa de Pós-Graduação em Letras da UFRGS, 2004. 121p. Dissertação de Mestrado.
7. Rocha Lima, Carlos Henrique da. Gramática Normativa da língua portuguesa. 21 ed. Rio de Janeiro: José Olympio, 1980.
8. Vale, Oto A. Expressões Cristalizadas do Português. Uma proposta de tipologia. Tese de doutorado. Araraquara- SP:UNESP, 2001.