

Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua

*Inacio Abdulkader**

Resumo: *O presente artigo apresenta os resultados de uma pesquisa de colocações verbais de have e go de interesse para o tradutor, que foi realizada utilizando-se corpora eletrônicos (Compara, Cobuild e WebCorp). Foram dois os tipos de buscas realizadas: corpus driven, e corpus based, e através delas foi possível o levantamento de um número de colocações que é bastante significativo frente ao que a lexicografia oferece para dois verbos tão freqüentes e de tal “dimensão gramatical” na língua. Ademais, procura-se chamar a atenção para a importância do poder expressivo da sintaxe dos strings de busca dos corpora que deve a nosso ver merecer uma maior atenção do pesquisador em corpora. No caso do Compara p. ex., essa sintaxe propiciou a construção de um “dicionário” que, muito provavelmente, incluiu a totalidade das ocorrências dessas colocações verbais no corpus. Além disso, o potencial dos corpora para pesquisas que vão além do aspecto lexical, envolvendo, p. ex., aspectos de interesse da Filosofia da Ciência e a da Linguagem é discutido e exemplificado.*

Palavras-chave: *corpora eletrônicos, colocações verbais, colocações*

Abstract: *This paper presents the results of a study on verbal collocations with “have” and “go”, which are of interest to translators working into Portuguese. Electronic corpora such as Compara,*

* O autor é engenheiro e aluno de pós graduação do programa de mestrado do Departamento de Letras Modernas da FFLCH-USP (kader@allnet.com.br).

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

Cobuild and WebCorp were the main tools used to obtain the results shown. Two types of searches were carried out, one “corpus driven” and one “corpus based”, which produced a significant number of collocations when compared to what is offered by the lexicographic aids for two verbs of such “grammatical dimension” within the language. We also discuss the expressive power of the syntax of the search string in the corpora, which proved to have a strength that, we believe, deserves closer attention. In the case of Compara, for instance, it permitted the construction of a “dictionary” that, probably, includes all occurrences of verbal collocations with “have” and “go” in that corpus at the time of search. We also point out, and try to exemplify, the potential of electronic corpora for research beyond the lexical level, including, for instance, aspects which are of interest to the Philosophy of Science and the Philosophy of Language.

Key words: *electronic corpora, verbal collocations, collocations*

1. Introdução

O presente artigo apresenta os resultados de uma pesquisa sobre colocações verbais de *have* e *go* com ênfase em aspectos de interesse do tradutor, e que utilizou como ferramenta básica os corpora eletrônicos. Essa pesquisa propiciou ademais ao autor seu primeiro contacto com os corpora, cuja existência até então desconhecia.

Assim, um objetivo adicional do artigo será o de descrever um pouco da experiência desse primeiro contacto com a pesquisa em corpus e deixar registradas algumas impressões sobre a ferramenta recém manuseada. Dessa forma, no que se segue haverá algum relato subjetivo e alguma digressão, além da usual apresentação objetiva de materiais, métodos e resultados.

O contexto é o de um aluno do primeiro semestre do CETRAD, o Curso de Especialização em Tradução com duração de quatro semestres que o Departamento de Línguas Modernas da FFLCH-USP oferece a graduados de todas as áreas que possuam o domínio requerido da língua de escolha.

¹ Souza, Beatriz, Inácio Abdulkader, Marcia Fiker e Stela Foley. “Colocações Verbais: os corpora como ferramenta para solução de problemas de tradução”, monografia submetida à disciplina de Lingüística Contrastiva II do CETRAD, DLM, FFLCH-USP, 2o semestre 2002.

Os resultados aqui apresentados são parte de um trabalho mais geral¹ enfocando colocações de vários outros verbos, elaborado por um grupo de quatro alunos para a avaliação de conclusão da disciplina “Linguística Contrastiva II” ministrada no segundo semestre de 2002. As três outras componentes do grupo cursavam àquela altura o último semestre e já tinham portanto tido algum contacto com a pesquisa em *corpus*.

Deve-se ademais registrar a gama das formações profissionais no grupo – uma física, uma psicóloga e uma bacharel em letras, além deste engenheiro – que é típica do ambiente multiprofissional que o CETRAD propicia, e que se revelou enriquecedora, dada a diversidade dos enfoques e a troca de experiências que proporcionou.

2. Corpus e Colocações

Cabe de início definir, talvez com um simplismo próprio de iniciante, o que é um *corpus* e o que são colocações.

Corpora eletrônicos são bancos de dados de textos. Ou seja, coleções de textos que são reunidos e catalogados de acordo com algum critério, armazenados de forma eletrônica, e que podem ser acessados ou apresentados de forma organizada ao usuário, conforme critérios que o usuário define.

“Critério” e “organização” são os conceitos chave. Num *corpus* que contivesse todas as peças de Shakespeare, o usuário poderia por exemplo solicitar que lhe fossem apresentadas todas as ocorrências da expressão “*whether it's nobler*” e mais, para uma primeira análise, as quinze palavras que a antecedem e lhe seguem em cada uma dessas ocorrências. Ou então solicitar da mesma forma as ocorrências de “*to be*”, caso em que o *corpus* apresentaria um número máximo de ocorrências, digamos quinhentas, estocasticamente colhidas dentre o total de instâncias de “*to be*” nas peças de Shakespeare. O texto mais amplo que inclui qualquer uma dessas ocorrências pode também ser exibido caso haja interesse.

Há portanto corpora de todos os tipos, organizados por autor, por assunto, por gênero, por época, corpora de registros da oralidade, corpora bilíngües e multilíngües, etc. A forma de apresentação pode também caracterizar o *corpus*. O *Compara*² por exemplo é um *corpus* literário, bilíngüe e paralelo. Inclui originais em português ou inglês e as respectivas traduções para a outra língua (várias traduções da mesma obra em alguns casos). Nele uma ocorrência é sempre apresentada de forma sinóptica com sua respectiva tradução (daí a denominação “corpus paralelo”).

² <http://www.portugues.mct.pt>

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

Um outro tipo de informação é a que se pode obter pela função *collocate sampler*, disponível em alguns corpora. Através dela as palavras que mais co-ocorrem com um determinado termo no corpus são listadas por ordem de frequência. Assim, um corpus que se componha de todas as edições de algum jornal brasileiro de grande circulação a partir de janeiro de 2003, com certeza indicará como *collocates* de alta frequência da palavra “Presidente” as palavras “Lula” e “Bush” (*hopefully* nessa ordem). E da palavra “Ministro” as palavras “José” e “Dirceu”. Como se vê pode-se extrair bem mais do que meras informações lingüísticas de um corpus.

Para o presente trabalho foram realizadas buscas no Compara e no Cobuild³, que é um corpus de textos os mais variados em inglês, já “fechado” – ou seja ao qual não se acrescentam novos textos – e que inclui cerca de quarenta milhões de palavras. Consultou-se também o Web Corp⁴ um *quasi-corpus* que na realidade usa toda a internet como seu banco de textos, permitindo, se necessário, algum controle da amplitude da busca (via buscador especificando-se, p. ex., google ou yahoo, via endereço, p. ex. PubMed, ou via domínio, p. ex., .br ou.uk).

Colocações, por outro lado, têm a ver com a convencionalidade da língua, com o “jeito como se fala”. Com a tendência de certas palavras de co-ocorrerem – aparecerem juntas com frequência na língua – e com o fato de que outras expressões, embora semanticamente corretas, são simplesmente descabidas na língua em uso. No que concerne ao tradutor, contrastes tais como “*have breakfast*” e “*tomar café [da manhã]*”, “*go nuts*” e “*ficar doido*”, mas não “*ir doido*” ou “*ir pras nozes*”.

Trata-se ademais de se poder reconhecer se um dado padrão lexical tem ou não um conteúdo semântico próprio e que difere, naquele caso específico, da “soma dos significados” dos componentes. Situações como a de, por exemplo, “*have us over a barrel*” que pode ser bem complicada, e não só para o tradutor.

Apresentado o contexto do pesquisador iniciante, e definidos o problema e a ferramenta, passemos a uma apresentação da metodologia e dos resultados.

3 Dificuldades iniciais e definição da metodologia.

A busca de colocações verbais de *have* e *go* desenvolveu-se em duas frentes. A primeira uma busca aberta, *corpus driven* – ou seja, direcionada em grande parte pelos próprios dados extraídos dos corpora conforme iam sendo obtidos – visou extrair do Compara o maior número possível de colocações verbais de *have* e *go* e suas respectivas traduções. A segunda, *corpus based* – ou seja, baseada na busca de específicas colocações verbais – realizou a busca de determinadas colocações nos

³ <http://www.titania.cobuild.collins.ac.uk/form.html>

⁴ <http://www.webcorp.org.ukwebcorp.html>

vários corpora visando a análise de aspectos semânticos particulares ou gerais que interessam ao tradutor, como os que mencionamos na seção 2 acima.

Circunstâncias específicas envolveram a procura e a definição da metodologia:

- 1) A tarefa foi inicialmente entendida, por quem estava tendo seu primeiro contacto com a pesquisa em corpora, como sendo a de “introduzir esses verbos nos corpora para obter suas colocações”.
- 2) A pesquisa se concentrou em dois dos verbos de maior frequência na língua inglesa, sendo um deles, aquele pelo qual se iniciou a pesquisa, um verbo auxiliar.
- 3) A pobreza da lexicografia, tão decantada pelos tradutores, é nesse caso ainda mais aguda. Embora existam dicionários de colocações (o mais recente o da Oxford de 2002) nos quais para cada verbete se lista as palavras que com ele mais co-ocorrem na língua, *have*, por ser verbo auxiliar, nem consta como verbete. Aparece citado ao longo de todo o dicionário nos verbetes cabíveis (p. ex., nos verbetes de *bath* e *drink*). E o verbete de *go* se restringe às mais óbvias colocações com preposições ou advérbios (*go well*, *go with*). Algo tão corriqueiro quanto *go home*, p. ex., aparece apenas no verbete *home*.
- 4) A pesquisa iniciou-se francamente *corpus driven* (*‘avant la lettre’* para quem a realizava) sendo que as palavras de busca inicialmente usadas nos corpus (Webcorp, Cobuild e Compara), foram as formas básicas de *‘to have’*. Assim, as primeiras buscas geraram uma quantidade não administrável de soluções, a imensa maioria delas com *have* na função de verbo auxiliar.

Essa situação gerou duas iniciativas que permitiram a definição de uma metodologia que conduziu aos resultados almejados:

- a) O suporte inestimável do pessoal de apoio do Compara possibilitou a construção de *strings* de busca que “filtraram” os resultados a favor de instâncias mais “léxicas” que propiciassem a ocorrência de colocações verbais. Com isso conseguiu-se extrair do Compara um considerável número de colocações verbais de *have* e *go*, algumas delas com várias ocorrências.
- b) Foi rapidamente reconhecida a necessidade de complementar-se as buscas do tipo acima com buscas de colocações específicas nos outros corpora, de forma a confirmar que alguns casos tratavam-se efetivamente de uma “colocação”. Além disso, essa busca de colocações específicas foi realizada sempre que havia interesse de se estudar mais a

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

fundo determinadas colocações. Assim, procurou-se reconhecer alguns pontos de natureza semântica importantes para o trabalho do tradutor.

4. Resultados das buscas *corpus driven*.

Foram **dois os tipos de strings** que o suporte do Compara nos proveu, de forma a contornarmos a “explosão” do *have* auxiliar.

O **primeiro tipo** pode ser exemplificado pelo *string* que transcrevemos a seguir, e que é o que consta da primeira linha do item b1 na **Tabela 1**, onde é denominado de **filtro 1**:

“have | had”[word!=”.*ed | been | made | brought | done”]{4,4}

TABELA 1
PESQUISA DE COLOCAÇÕES VERBAIS DE “HAVE” NOS CORPORA

a) **Webcorp** (sem filtros*): “have”, “have - *ed”, “had”: 1902 *concordances*.

11 colocações verbais
(6 coincidentes com obtidas no COMPARA)

0,57%

* a não ser por “have-*.ed”.

a) **Compara** (com filtros): 1967 *solutions*,

32 colocações verbais

1,62%

b1) filtro 1 (dois casos):

“have | had”[word!=”.*ed | been | made | brought | done”]{4,4},

1787 *solutions*, 500 *shown*

“having | has” [word!=”.*ed | been | made | brought | done”]{4,4},

467 *solutions*

b2) filtro 1 a:

“have | had”[word!=”.*ed | been | made | brought | done | thought | se

en | broken | left | bought | sent | had | given | to | set”]{4,4}, 1200

solutions, 500 *shown*

b3) filtro 2:

“have | had” “a | the | my | some | to”, 608 *solutions*, 500 *shown*

TOTAL: 38 COLOCAÇÕES VERBAIS DE “HAVE”

Ele é “traduzido” como: “busque as formas “have” ou “had” seguidas por (exatamente) quatro ocorrências de formas que não terminam em “ed”, e que não são as formas “been”, “made”, “brought”, “done””.

O *string* na segunda linha do item b1 na **Tabela 1** é igual ao acima com a diferença de que é aplicado para as formas “having” e “has”. E o “**filtro 1 a)**”, no mesmo Apêndice, é também um *string* deste primeiro tipo, apenas mais restritivo, pois “veta”, que a “have” e “had” se siga um maior número de participios irregulares.

O **segundo tipo** de *string* tem a sintaxe do **filtro 2** que consta no item b3 da **Tabela 1**:

“have | had” “a | the | my | some | to”

Essa sintaxe busca os casos em que é provável que um participio não siga a expressão buscada, no caso, “have a”, “have some”, “have the”.

Esses “filtros” tornaram a busca no Compara manejável e permitiram obter desse corpus 32 tipos de colocações verbais de *have* (**Tabela 1**), num total de 56 ocorrências. A **Tabela 1** mostra também os números bem mais reduzidos de colocações obtidos nas buscas (sem filtros) de *have* no corpus do Cobuild, e também no Webcorp, sendo que esta última busca tornou-se manejável apenas na medida em que os resultados do Compara nos alentaram a enfrentá-la. Realizamos também uma busca com o *collocate sampler* do Cobuild que nos propiciou obter apenas um novo tipo de colocação verbal.

Ao fim, totalizamos 38 diferentes colocações verbais de *have*, listadas na **Tabela 2**. Esse número é significativo frente ao que os dicionários e correlatos oferecem. Das colocações levantadas através dessa literatura, que são quantificadas na **Tabela 3**, nem todas são verbais. Normalmente aparecem sob outros títulos (p. ex. “Have” em Swan: 2002), visando ilustrar outros aspectos e não o fato de se tratarem de colocações verbais. E são, geralmente, colocações bem conhecidas.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

TABELA 2
Colocações de Verbais "Have" Obtidas

Have a baby	Have a termination
Have a bath**	Have an affair
Have a bite	Have an argument
Have a child	Have an operation
Have a degree	Have breakfast
Have a go	Have eyes for
Have a good laugh	Have food
Have a haircut	Have fun
Have a head start*	Have in mind
Have a look	Have priority
Have a nice time*	Have qualms***
Have a nightmare	Have second thoughts
Have a part*	Have sex
Have a role*	Have some sleep
Have a row	Have the curse
Have a say	Have the heart
Have a scotch	Have the nerve
Have a seat	Have therapy
Have a temperature	Have trouble

(*) só no Webcorp

(**) no Webcorp e no Cobuild

(***) do collocata sampler do Cobuild

todas as demais colocações são do COMPARA

TABELA 3
COLOCAÇÕES DE "HAVE" EM MANUAIS, DICIONÁRIOS E GRAMÁTICAS

a) Oxford Collocations Dictionary (2002): nenhuma colocação (i.e., não tem verbete para "have").

b) Cambridge Grammar of the English Language (2002): 6 colocações verbais, 4 colocações "não-verbais". ("Collocations" não consta do índice remissivo de conceitos).

c) Bloomsbury Good Word Guide (1990): nenhuma colocação

d) Swan, Practical English Use (1995): 14 "tipos" de colocações verbais de "have"—i.e, p. ex. "have lunch" e "have dinner" são consideradas do mesmo "tipo". (Citadas no parágrafo sobre "have: actions" e não sob "collocations").

e) Tagnin, Expressões Idiomáticas e Convencionais (1989): 4 colocações verbais, 3 "não-verbais".

f) Collins Dictionary of Phrasal Verbs and their Idioms (1974): nenhuma colocação verbal, 6 "phrasal verbs".

A busca de colocações verbais de *go* no Compara seguiu procedimentos muito semelhantes aos adotados para a busca dos de *have*. Não sendo *go* um verbo auxiliar, não ocorreu aqui nenhum fenômeno “explosivo” como no caso de *have*. Em nenhuma situação o total de ocorrências ultrapassou as quinhentas que são o número máximo de exibições do Compara. Como se vê na **Tabela 4**, para as formas “*go*” e “*goes*” não foi necessário utilizar nenhum filtro.

TABELA 4
PESQUISA DE COLOCAÇÕES VERBAIS DE “GO” NOS CORPORA

a) Compara total de:
22 colocações 3,26%
“ <i>go</i> ” e “ <i>goes</i> ” (sem filtros): 310+58 <i>solutions</i>
“ <i>went gone</i> ” e “ <i>going</i> ” (com os filtros abaixo): 202 + 104 <i>solutions</i>
"went gone" [word!=".*ed to away somewhere anywhere nowhere on off into"]{4,4}
e
"going" [word!=".*ed to away somewhere anywhere nowhere on off into"]{4,4}

b) Cobuild (sem filtros): 80 <i>concordances</i> , 2 colocações não coincidentes com o Compara.

c) Cobuild <i>collocate sampler</i> : 6 colocações (não coincidentes com as acima)
--

d) Webcorp (busca específica): 1 colocação
--

TOTAL: 31 COLOCAÇÕES DE “GO”

O filtro utilizado para “*went|gone*” e “*going*”, transcrito abaixo para o caso de *went|gone* (item a da **Tabela 4**), é um *string* do primeiro tipo acima descrito. “*went|gone*”[word!=".*ed|to|away|somewhere|anywhere|nowhere|on|off|into"]{4,4}

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

Sua aplicação não foi ditada por qualquer “explosão” de resultados, que, como vimos, não ocorreu nas buscas de *go*. Ela visou apenas facilitar o *sorting*, evitando alguns advérbios de lugar e movimento, *ing-words* (via “*on*”, “*off*” e “*into*”), e participios (via “**ed*”).

Do total das 31 colocações verbais de *go* levantadas (vide **Tabela 5**), 22 são do *Compara* totalizando 40 ocorrências de colocações verbais de *go* nesse corpus. Não sendo *go* um verbo auxiliar – portanto mais “léxico” ou menos “gramatical” que *have* – o *collocate sampler* do Cobuild forneceu 6 interessantes colocações contra apenas 1 de *have*. Da **Tabela 6**, vê-se que o número de colocações verbais por nós obtidas é, no caso de *go*, ainda mais significativo, frente ao (quase nada) que se consegue extrair da literatura.

TABELA 5
Colocações Verbais de “Go” Obtidas

Go amiss*	Go in late	Go public***
Go astray*	Go into detail	Go quiet
Go awry*	Go mad	Go to any trouble
Go bald	Go nuts	Go to bed
Go bankrupt	Go off beat**	Go to perdition
Go berserk*	Go off the rails	Go to prison
Go blank	Go on stage	Go to sea
Go crazy	Go on the run***	Go wide
Go haywire*	Go out of (their) way	Go wild
Go his own way	Go overboard*	Go wrong
		Go years back

* do *collocate sampler* do Cobuild
 **busca específica no *Webcorp*
 *** do Cobuild
 todas as demais colocações são do *COMPARA*

TABELA 6
COLOCAÇÕES DE “GO” EM MANUAIS, DICIONÁRIOS E GRAMÁTICAS

-
- a) **Oxford Collocations Dictionary (2002):** nenhuma colocação verbal, 15 colocações adverbiais e phrasal verbs .
- b) **Cambridge Grammar of the English Language (2002):** nenhuma colocação verbal, 2 colocações “não-verbais”. (“Collocations” não consta do índice remissivo de conceitos).
- c) **Bloomsbury Good Word Guide (1990):** nenhuma colocação
- d) **Swan, Practical English Use (1995):** nenhuma colocação.
- e) **Tagnin, Expressões Idiomáticas e Convencionais (1989):** 4 colocações verbais, 1 “não-verbal”.
- f) **Collins Dictionary of Phrasal Verbs and their Idioms (1974):** nenhuma colocação verbal, 37 “phrasal verbs”.
-

Ademais, é bem provável que tenhamos levantado **todos** os tipos de colocações verbais de *go* existentes no Compara à data das buscas, e a **quase totalidade** dos de *have*. Isso demonstra o considerável poder expressivo da sintaxe de busca do Compara, que merece ser comentado com mais detalhe.

4.1 O potencial dos “filtros”.

À parte os “filtros” terem sido fundamentais para a viabilização de nossas buscas por colocações verbais de *have*, sua utilização indicou que a sintaxe dos *search strings* do Compara é suficientemente forte (ou seja, consegue exprimir suficientes conteúdos semânticos) para merecer uma maior atenção do tradutor, de forma a que o potencial de extração de dados desse corpus possa ser devidamente explorado. No presente caso ao menos, e que certamente é típico de uma série de outras buscas⁵, tudo indica que os “filtros” compensaram muito

⁵ Em geral as expressões cuja sintaxe por um lado deve incluir palavras ou formas de alta frequência na língua, e por outro deve excluir outras palavras ou formas também muito frequentes na língua. Uma solução que consegui “filtrar” as formas em que um verbo funciona como auxiliar, servirá para uma série de outras situações que requerem um menor *screening*.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

riamente o fato de o Compará não ser um corpus anotado⁶. E que podem chegar a compensá-lo plenamente, em termos de propiciar a extração da totalidade das colocações existentes nesse corpus.

Recorde-se de início que é através de um *sorting* das soluções exibidas pelos corpora – i. e. do exame de cada solução – que se localiza as colocações existentes no corpus. E que para todos os três corpora apenas uma pequena porcentagem das soluções exibidas (entre 0,5 e 3,3%, vide Tabelas 1 e 4) são colocações. Graças à utilização dos filtros, as porcentagens mais altas são as do Compará.

No que respeita especificamente ao Compará e seus filtros, note-se que em apenas três de nossas buscas (que são, todas as três, das formas “*have|had*”) o número de soluções ultrapassou as quinhentas, que é o número máximo de soluções que podem ser apresentadas numa exibição. Isso determinou para esses casos uma escolha randômica pelo sistema das soluções que foram exibidas. Em todas as outras buscas (que incluem a totalidade das buscas de *gθ*) o número de soluções foi inferior a 500, o que indica que teremos levantado nesses casos, salvo falta de atenção ou erro de julgamento, todas as colocações verbais das respectivas *wordforms* no Compará.

Os três casos com mais de 500 soluções foram: uma busca com o filtro 1 que obteve 1787 soluções, outra com o filtro 1a) que obteve 1200, e a terceira com o filtro 2 obtendo 608 soluções.

Por um lado, isso indica que devem (ou podem) existir no Compará colocações de *have* que não nos foram exibidas. Por outro lado, um cálculo de probabilidade bem simples (s^3 de V para curva normal) indica que mais uma reiteração de exibição da busca do filtro 2, duas mais de exibições do filtro 1 a) e três do filtro 1, propiciarão a exibição da totalidade das colocações verbais de *have* no Compará com mais de 99,9% de probabilidade.

Os casos que parecem ser mais críticos são os dos filtros 1 a) e 1, em que o número de soluções exibidas é bem inferior (respectivamente, cerca de duas e três vezes) ao de soluções encontradas. Mas como o filtro 1 a) gera subconjuntos das soluções do filtro 1, as exibições do filtro 1 a) constituem uma espécie de reiteração parcial das exibições do filtro 1. Isso reduz para duas por filtro o número de reiterações requeridas para que se consiga a exibição – estatística – do total de colocações constantes nas soluções de 1 a) e 1. Já no caso do filtro 2,

⁶ Corpora anotados são aqueles cuja base de dados contém, além dos textos, comentários e anotações sobre o conteúdo desses textos. Ideal para o caso do presente estudo seria um corpus que assinalasse todas as colocações verbais que ocorrem em seus textos.

o número de soluções por exibição é bem próximo do de soluções totais (500 de 608), de forma que uma reiteração, como vimos, já é suficiente.

Há ainda um recurso de outra natureza: a melhora da **eficiência** das exibições pelo aumento da porcentagem de soluções relevantes em cada exibição de 500 soluções. Isso pode ser obtido pelo aumento da lista de palavras restritivas nos filtros de primeiro tipo. Para tanto, serve como orientação o resultado (ou melhor, as soluções descartáveis) de exibições anteriores, menos restritivas.

Assim, através de recursos como os acima descritos e de outros que se venha a desenvolver, pode-se evitar a explosão do número de soluções irrelevantes que tanto desalenta (como desalentou o iniciante seu primeiro input, *naive*, no Compara).

E esse refinamento do *string* de busca tornar-se-á cada vez mais importante com o crescimento da base de textos do Compara. Ele será fundamental, por exemplo, para que no futuro o trabalho envolvido no *sorting* de soluções exaustivas do corpus, mantenha-se, como é hoje, dentro do que é razoavelmente manejável.

A lição é clara mesmo para quem apenas se inicia na busca em corpus: a sintaxe de busca dos corpora, em especial para o tradutor a do Compara, tem um poder de expressão semântica dos requisitos de busca que merece ser mais bem explorado.

É importante que se aprofunde o estudo da sintaxe de busca dos corpora, para que o *string* de busca vá além de uma passiva introdução da palavra, morfema, ou expressão no campo de *search*, na esperança de que as exibições surjam, numerosas e relevantes.

Com isso o potencial da ferramenta corpus passa a ser explorado de forma mais plena e eficaz.

4.2 Alguns resultados do Compara

Abaixo, nos restringimos a apresentar cinco casos de colocações verbais de *have* e *go* menos usuais, obtidas no Compara através das buscas acima descritas. Algumas colocações “não verbais”, ou seja *phrasal verbs*, expressões idiomáticas ou fixas obtidas nessas buscas e que consideramos ser de interesse para o tradutor, são apresentadas no **Apêndice 1**.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

EBDL1T1 (1660):	<i>At that moment I was surprised to hear Lewis Parker saying from the set, «Well, if she is pregnant she'll have to have a termination,» and Debbie replying, «I suppose you think that will solve everything.»</i>	Nesse momento surpreendeu-me ouvir as palavras do Lewis Parker vindas do cenário: «Bem, se ela está grávida, tem de fazer um aborto.» E a Debbie responder-lhe: «Deves achar que isso resolve tudo.»
EBDL1T2 (1945):	<i>I had a go at The Concept of Dread this evening - thought I'd start with the title that seemed most obviously relevant to me - but it was a great disappointment.</i>	Comecei a ler <i>O conceito de angústia</i> hoje à noite - achei que seria bom começar pelo título que mais combinava comigo --, mas foi uma grande decepção.
EBIB1(949):	<i>Avoir ses ours? , to have one's bears, means 'to have the curse' (presumably because at such times the woman is supposed to behave like a bear with a sore head) .</i>	<i>Avoir ses ours</i> significa estar com o período (possivelmente porque se pensa que nessas alturas a mulher tem o comportamento de um urso com mau perder).
EBDL1T2 (1973):	<i>I would say that learning to know dread is an adventure which ever man has to affront if he would not go to perdition either by not having known dread or by sinking under it.</i>	Eu posso dizer que aprender a conhecer a angústia é uma aventura com que cada homem tem de confrontar-se se não quiser cair na perdição ou por não conhecer a angústia ou por se afundar nela.
EBJT1(776):	<i>Some hope, he thought now, putting cheese between the bread slices, some hope, with costs going up all the time, Gareth on £13,000 plus his house, cows needing several hundred pounds ' worth of bought-in feed a year each, interest rates... and now this cold wet spring would mean the maize would go in late.</i>	Alguma esperança, pensava ele pondo queijo entre as fatias de pão, alguma esperança, com os preços a subirem constantemente, com Gareth a custar 13 000 libras mais a casa, com as vacas a precisarem todos os anos de rações em que despendia várias centenas de libras, as taxas de juro... e agora esta Primavera fria e chuvosa, o que significava que o trigo se iria atrasar .

5 Resultados das buscas corpus based.

Além da acima descrita busca aberta no *Compara*, foram realizadas, como já mencionado, buscas específicas de algumas colocações verbais de *have* e *go* no *Cobuild* e no *Web Corp*. Essas colocações (inclusive as do *Compara*) estão listadas nas **Tabelas 7 e 7A**, junto com o número de ocorrências obtidas para cada uma delas em cada corpus.

Nos casos das colocações menos comuns a busca visou confirmar que a expressão encontrada no *Compara* tratava-se mesmo de uma colocação, com o uso e conteúdo semântico nele (*Compara*) apontados sendo efetivamente empregados em outros exemplos.

No caso de colocações mais frequentes, o objetivo foi o de aprofundar o estudo da colocação em si e da natureza das colocações verbais em geral, procurando reconhecer exemplos em que o(s) significado(s) e as diversas formas de emprego fossem evidenciados. A seguir apresentamos dois exemplos mais signi-

Tabela 7

Busca Específica de colocações verbais de “*Have*” nos Corpora

Have a bath (132)
Have a drink (89+3)
Have a look (60+8)
Have a nice time (97)
Have a role (108)
Have an argument (133)
Have fun (42)
Have the heart (132+1)
Have a go(62)
Have a temperature (20)
Have a termination (118)
Have the curse (2+1)
Webcorp + COMPARA

TABELA 7 A

Busca Específica de colocações verbais de “*Go*” nos Corpora

Go amiss 10**
Go awry 9**
Go bald 183*
Go berserk 157*
Go in late 40*
Go mad 72*
Go off beat 17*
Go off the rails 106*
Go on the run 118* + 1**
Go out of their way 134*
Go overboard 14**
Go public 119* + 1**
Go to perdition 126*
Go to sea 244*
Go years back 34*
*Webcorp **Cobuild

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

ficativos com as observações de interesse para o tradutor que julgamos pertinentes para cada caso.

(1) http://www.download2me.com/jokes/joke_content/dart_board.htm
The Model Lodger

Doris and Fred had started their retirement years and decided to raise some extra cash by advertising for a lodger in their terrace house. After a few days, a young attractive woman applied for the room [...]

“There’s just one problem,” explained the model. “Because of my job, I have to **have a bath** every night, and I notice you don’t **have a bath**.”

Neste primeiro exemplo temos, na mesma sentença, duas ocorrências da expressão “have a bath”. Os conteúdos semânticos são, no entanto, totalmente distintos. O primeiro é o da colocação verbal “tomar banho” (que denominaremos doravante “sentido **colocativo**” correspondendo a “*collocate sense*” em inglês). No segundo, claramente, cada palavra retém um conteúdo semântico próprio e usual: “têm (= dispõem de uma) banheira”. É o que denominaremos “sentido **livre**” ou “**comum**” da expressão – correspondendo a seu “*plain sense*”, em inglês. É interessante notar que, se não estiver especificamente “em busca” de ocorrências da expressão, o leitor/tradutor (como o próprio autor da sentença) nem “repara” que está lendo/traduzindo (/escrevendo) duas vezes a mesma frase sintática.

(2) <http://www.mtholyoke.edu/~ebarnes/python/argument-clinic.htm>
Monty Python: Argument Clinic

*Man: Ah. I'd like to **have an argument**, please.*

Receptionist: Certainly sir. Have you been here before?

[...]

Man: Oh look, this isn't an argument.

Mr Vibrating: Yes it is.

Man: No it isn't. It's just contradiction.

Mr Vibrating: No it isn't.

Man: It is! [...] : An argument isn't just contradiction

Mr Vibrating: It can be.

Man: No it can't. An argument is a connected series of statements intended to establish a proposition.[...]

Aqui um *sketch* humorístico, a situação que tanta dificuldade traz para o tradutor. Como traduzir “*have an argument*”? “Comprar uma briga” é a tradução óbvia de início, mas o tradutor logo se deparará com o “intraduzível” ao chegar na

“briga” sobre “*what is an argument*”. O importante neste caso é notar que, ao contrário do que acontece no exemplo anterior, a existência de dois sentidos para a mesma frase sintática – neste caso **concomitantes** na **única** instância da frase que surge no texto acima – não só se faz notar mas é inclusive ressaltada intencionalmente. O autor joga o sentido **colocativo** contra o sentido **livre** para criar a situação de humor básica para o *sketch*.

Exemplos poderiam suceder-se: “*go off beat*” claramente significando “sair do ritmo” num caso; claramente significando “mudar de assunto” em outro; ou ficando no “meio” dos dois, num site musical. “*Go out of their way*” significando “fazem de tudo”; ou então criando uma situação entre o **livre** “desviar do caminho” e o **colocativo** “fazer de tudo”, que soa de péssimo gosto num site que busca conscientizar os motoristas a respeitarem os ciclistas. A frase de um site de doação de órgãos: “*Do you have the heart to be an organ donor...*”, cuja tradução, é claro, teria que seguir o sentido **colocativo** e perder a alusão: “Você tem coragem de ser um doador de órgãos”. E que no entanto “clama” por uma solução de tradução (não encontrada), para que a alusão não se perca.

A lição aqui é a de que sentido **livre** e sentido **colocativo** são os dois extremos de um contínuo e não um “zero ou um”. Assim, o requisito para a correta tradução de uma colocação verbal muitas vezes vai além de saber-se que ela existe e conhecer o seu sentido **colocativo**.

Em certas situações, o autor joga com ambos os sentidos, **colocativo** e **livre**, e aí se encara o intraduzível. Noutras, é necessária uma análise para que se decida qual dos dois sentidos adotar, em qual “ponto do contínuo” se situa aquela determinada acepção. Essa decisão nem sempre é simples, como se vê no exemplo 7 do Apêndice 2 e no caso exposto em 5.1 abaixo, onde algo mais do que a mera tradução está em jogo.

Ainda no que respeita à “definição do ponto do contínuo”, foi possível – e interessante – reconhecer para o caso de pelo menos uma colocação (*have the heart*) a existência de algo como um “índice sintático da posição no contínuo”, que pode orientar o tradutor.

A **preposição** que se segue a “*have the heart*” parece indicar essa posição no contínuo de acepções. Seguida de “*to*” (ou quando situada no fim de uma sentença, caso em que o “*to*” é elíptico), a colocação claramente traduz-se por “coragem [a despeito de tudo]”, como nos exemplos 1 e 3 do Apêndice 2. Quando seguida de “*for*” como no ex. 4, a tradução ainda é “coragem” mas a coragem vem acompanhada de algo como “a predisposição, o talento” que abranda a “ousadia da coragem”. Esse é o caso de, por exemplo, “*I have the heart for the job*”.

Finalmente, quando seguida de *of* a acepção já se situa bem mais próxima do outro extremo do contínuo e a tradução será, normalmente, “ter o coração” (na acepção de “ter o espírito”, vide exemplo 2 do Apêndice 2).

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

5.1 Referência e tradução

Isso posto, a busca em corpus, ao que parece, pode às vezes render mais do que se espera. A seguir examinamos uma outra ocorrência de *have a heart* no Web Corp, que merece uma reflexão mais atenta.

Consideremos o seguinte texto, manchete e trecho de uma notícia do *The Independent*:

(3) <http://www.balkanpeace.org/hed/archive/july00/hed402.shtml>
The Independent (UK), July 30 2000
By Phil Rees in Podgorica

We have the heart for battle, says Montenegrin trained by SAS [...] *The Special Police has a fierce reputation in Montenegro its gung-ho approach seemingly unsettling the SAS. "They thought we were crazy. When two of us banged into a house and started shooting into walls, bullets were flying around and they said 'Oh, it's a real gun, real bullets? You're crazy guys, you don't have protection'. But **we have a heart, we don't have protection but we have a heart. A big heart.**"*

Nesse texto jornalístico não há a princípio maior dificuldade para a tradução da colocação na manchete. Ela diz: "Temos coragem para a batalha", ou, numa tradução menos literal que talvez soe melhor em português, "Somos destemidos no combate". Nem há dificuldade na frase do combatente montenegrino que consta do corpo da notícia. Sua tradução seria algo como: "...temos coração, não temos proteção mas temos coração. Um grande coração".

A pergunta (de quem traduz para um redator que vai decidir a manchete em português) é se a manchete em inglês reflete de fato a frase do entrevistado, frase que o jornalista Phil Rees toma como o aspecto mais marcante da notícia a ponto de utilizá-la em sua manchete. E a utiliza na manchete como quem citasse *ipsis verbis* o entrevistado, o que, já se vê, não é o caso. Por um lado, o entrevistado não usa a palavra "*battle*" – mas, é claro, o contexto autoriza o jornalista a considerá-la em "elipse" na frase. Por outro lado, o entrevistado não diz "*we have the heart*", mas, "*we have a heart*". Estaria ele se referindo (tão somente) à coragem ou a "algo maior"?

Posta em outros termos, a questão é a de se estabelecer se a informação que foi captada pela grande maioria⁷ de leitores do *The Independent* – aqueles que apenas passaram os olhos pela manchete e não leram o corpo da notícia – corresponde ou não à realidade do que expressou o entrevistado.

⁷ Vide Lonardoni (1999, p.112), citando o *Novo Manual de Redação* da Folha de São Paulo. Em seu artigo Lonardoni apresenta também um modelo de T. Van Dijk para a superestrutura do discurso da notícia na imprensa escrita. Esse modelo aborda, entre outros temas, o que de comum deve haver (ou o leitor espera que haja) entre

A resposta a essa questão requer uma digressão.

* * *

W. V. Quine (1908-2000), um dos mais importantes filósofos norte-americanos do pós-guerra, exerceu grande influência na corrente analítica do pensamento recente em áreas como a lógica, a filosofia da linguagem, a epistemologia e a metafísica.

Sua doutrina mais famosa, a da “indeterminação da tradução radical”, foi formulada no âmbito da epistemologia – e não no da teoria da tradução⁸ – como refutação a certas posições do empiricismo⁹.

Com essa doutrina Quine visava estabelecer o “naturalismo”, o ponto de vista de que a Filosofia deveria ser desenvolvida num “contínuo” com as Ciências sem qualquer status radicalmente distinto para a Matemática ou a Lógica, e sem o apelo a certas “verdades analíticas”, ou seja, verdades que são verificáveis em virtude do “significado” das respectivas proposições.

Não há “significados” que tenham tal poder ou tal “transcendência”, diz Quine, na medida em que não existem referentes que possam ser absolutamente determinados.

O que cabe aqui notar é que para apresentar seu argumento acerca da indeterminação dos **referentes** de uma forma “menos abstrata e mais realística” (QUINE: 1960, 27), Quine viu-se levado a lançar mão da **tradução** como **exemplo**. E a descrever uma situação hipotética de “tradução radical, i.e., a tradução

os conteúdos da manchete e do corpo da notícia (em especial o conteúdo da “lide” o trecho inicial do corpo da notícia). Algo com implicações para a tradução de textos jornalísticos e mesmo para a própria ética do jornalismo.

⁸ Note-se no entanto que o argumento de Quine, basicamente um alerta contra a sinonímia, influenciou teóricos da tradução como I. A. Richards, Anton Popovic e mais recentemente Anne Mette Hjort. Dada a estatura de Quine, o nome adotado para a doutrina, e a forma de apresentação de seu argumento, esse tipo de influência seria de se esperar. Veja-se Gentzler: 1993, pp. 15,81,197.

Cabe ainda notar que Quine lecionou no Brasil, na FFCL-USP, por três anos no início da década de 1940. Publicou em português um texto com conceitos que se tornaram seminais no desenvolvimento da Lógica, *O Sentido da Nova Lógica* (Editora Martins, 1944).

⁹ Posições que valorizavam o significado das proposições como algo que define o tipo de experiências (científicas) que poderiam vir ou não a contar como evidência, contra ou a favor, do valor de verdade dessas proposições.

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

da língua de um povo até então intocado” (id., 28) com o qual um lingüista pela primeira vez se depara, no momento exato em que um coelho passa chispando na frente de um nativo, o qual, ato contínuo, grita “gavagai”.

À moda dos filósofos, Quine passa boa parte de “Translation and Meaning”, o hoje famoso segundo capítulo de sua obra maior, *Word and Object*, a destruir qualquer esperança de que o lingüista possa concluir que a tradução de “gavagai” é “coelho” e não outra coisa qualquer como por exemplo “meros estágios, ou breves segmentos temporais, de coelhos” (ibid., 51). É impossível, argumenta Quine, que se decida algo de definitivo acerca de qual é, na língua do “povo intocado”, o referente de “gavagai”.

É preciso ressaltar novamente que Quine não está argumentando contra a possibilidade da tradução – que na realidade do dia-a-dia é “não-radical” e totalmente factível e corriqueira (ibid., 28) – mas contra uma certa, digamos, reificação ou “transcendência” do significado. Para isso, Quine busca demonstrar a indeterminação de **todo tipo** de referência. Nada contra a tradução em si.

Por outro lado, é significativo que uma situação extrema da tradução – ainda que idealizada – tenha sido a escolhida para a explicitação do argumento, e que isso o tenha marcado a ponto de ter-se adotado para a doutrina uma denominação algo imprecisa: “indeterminação da **tradução radical**”¹⁰. Melhor teria sido talvez – inclusive para o bem da Tradução – a denominação mais precisa “doutrina da indeterminação da **referência**”.

* * *

Pois bem, a matéria do *The Independent* que aflorou no Web Corp traz uma espécie de gavagai real. A dúvida do tradutor é “quineana”: não se trata (só) de encontrar a tradução, mas de se determinar qual é o referente, que **uso** o entrevistado estará fazendo da expressão “*have a heart*”.

- 1) O inglês do entrevistado será apenas sofrível, ele tem conhecimento de uma certa colocação com *have* e *heart* que significa “coragem”, e ao utilizá-la confundiu o artigo?
- 2) Ou “*a heart*” aqui é “mais que coragem”. É “um coração”, “uma razão maior”, “energia, entusiasmo, ardor”, dos quais a “coragem ousada” é apenas uma das conseqüências? (Note-se que *heart*, isoladamente, pode

¹⁰ Ademais, o fato de ser mais freqüentemente adotada uma denominação abreviada, “indeterminação da tradução”, que omite tratar-se da “tradução radical” tal como definida por Quine, facilita a interpretação apressada e superficial de que “**toda** tradução é, a rigor, impossível

ter a acepção específica de “coragem”, como por exemplo em “*I plucked whatever heart I could and said...*” (NSOD), mas **não** quando precedido pelo artigo indefinido “*a*” como na frase do entrevistado. Por outro lado o “*for*” que se segue à colocação na manchete não ajuda. Como vimos acima ele “abranda a coragem” que no caso é no mínimo “ousada”).

No primeiro caso a manchete estaria “correta” (i.e., refletiria a realidade do que disse o entrevistado). No segundo caso, o *plain sense* (com *heart* num sentido figurado) terá sido tomado, equivocadamente pelo jornalista, pelo *collocate sense*.

E há ainda a possibilidade de tanto a acepção de “*have a heart*” para o entrevistado, quanto a de *have the heart* para o jornalista (não obstante o *for* que se segue à colocação), ficarem neste caso “no meio do caminho”, entre a “coragem” e a “razão maior”, entre o *collocate sense* e o *plain sense* (figurado) de “*have the heart*”.

Enfim, difícil de decidir. E assim, em respeito a Quine, deixamos a questão indecisa. Já o tradutor no mundo real, premido é claro pelo *deadline*, terá tomado uma decisão e produzido a tradução.

Seja como for, eis aqui um exemplo de que os corpora como ferramenta não se prestam apenas ao estudo da língua em uso. Deles se pode extrair também aspectos referentes à epistemologia e às filosofias da ciência e da linguagem.

Mas ainda há mais com respeito aos corpora.

6. Carving Language

Ao estudar a estrutura e função das teorias científicas e de seus conceitos, Carl Hempel (1905-1997), um dos líderes do “empiricismo lógico” na filosofia da ciência, define como “relevância sistemática (*systematic import*) de um conceito” o fato de que ele se preste à formulação de princípios teóricos que reflitam as uniformidades do objeto de estudo (HEMPEL:1965, 146). Esse requisito, diz-nos Hempel, aplica-se também no caso das classificações científicas e está por traz da distinção que intuitivamente se faz entre **classificações “naturais” e “artificiais”**.

Por exemplo, os dois conjuntos de características sexuais primárias que determinam a divisão dos seres humanos entre homens e mulheres estão, cada um desses conjuntos, associados através de leis ou de correlações estatísticas a uma larga variedade de outras características de várias ordens e que lhes são concomitantes: características físicas, fisiológicas, psicológicas, etc. Isso contrasta com uma classificação que, por exemplo, dividisse os seres humanos em “até

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

1,70 m de estatura” e “acima de 1,70m” (classificação esta que pode até ser útil em casos específicos).

Nesse contexto Hempel recupera uma expressão que remonta a Platão e que se tornaria lapidar para designar os conceitos que fazem avançar as ciências. Diz ele do primeiro exemplo de classificação acima: “*It is understandable that a classification of this sort should be viewed as somehow having objective existence in nature, as ‘carving nature at its joints’*”.

Essa expressão vem à mente quando se reflete de forma mais ampla acerca do que ocorre na pesquisa em corpora. Foi fascinante constatar o quanto essa pesquisa é um “mergulho” na língua viva. Através dela uma série de conceitos afloram, são inferidos de maneira natural, automática. E a forma como esses conceitos se articulam transparece.

Pode-se de certa forma dizer que, ao utilizar os corpora para prospectar a língua, vivencia-se a realidade de alguns conceitos como “*somehow having objective existence in language, as ‘carving language at its joints’*”.

Por exemplo, a dimensão léxico-gramatical das palavras e o fato de ela constituir um contínuo, se revelou através da grande efetividade do *collocate sampler* quando aplicado a *go* para buscar colocações, em Comparação com sua baixa efetividade para *have*. E no fato de para *have* (mas não para *go*) terem sido necessários *strings* na forma de “filtros” para que a busca de colocações se tornasse manejável. Assim, essa gradação das palavras de “gramaticais” a “léxicas”, com toda uma gama intermediária de palavras que como *have* e *go* exercem ambas funções em diferentes graus, “salta aos olhos” via os corpora. Isso demonstra que os conceitos de “dimensão léxica” e “dimensão gramatical” têm *systematic import*, indica que “*they carve language at a joint*”.

O exame de colocações específicas nos corpora revelou um outro contínuo, desta feita o da “acepção” da expressão, cujos extremos denominamos sentido “livre” ou “comum” (*plain sense*) e sentido “colocativo” (*collocate sense*). A necessidade de se avaliar o “ponto” exato em que se “está” nesse contínuo, e a possibilidade de que exista alusão de um *sense* ao outro constitui uma realidade que precisa ser levada em conta pelo tradutor. Ou seja, o fato de que os sentidos livre e colocativo não são um “zero ou um” é um outro *joint of language carved via corpora*. Essa constatação corresponde – para o caso específico das colocações que estudamos – ao que, em termos mais gerais, Baker (1992, pp. 205-206) denomina *instantial meaning* de uma palavra ou expressão num texto, que é específico de cada texto, e que o próprio ambiente textual determina.

Nesse sentido, um aspecto que discutimos para o caso de *have the heart* na seção 5 acima é interessante. Como se recorda foi possível utilizar a preposição que se segue a essa colocação – “*to*”, “*for*” ou “*of*” – como uma espécie de

“índice da posição daquela ocorrência no contínuo de acepções da colocação”. Isso significa que, graças aos corpora, identificou-se de uma forma bastante objetiva “um tijolo” – por mínimo que seja – da construção dos vários *instantial meanings* de “*have the heart*”. Tem-se aí um pouco de *instantial meaning* “*in the making*”. Um pouco da semântica emergindo e se **articulando**, a partir da sintaxe.

Por outro lado os corpora podem indicar que certas classificações tradicionais nas gramáticas não têm base na forma como a língua se articula. Isso parece ser o caso da distinção entre colocações verbais e não verbais, colocações verbais e *phrasal verbs*, e mesmo entre colocações, expressões fixas e expressões idiomáticas.

Aparentemente trata-se sempre do mesmo fenômeno básico: a co-ocorrência freqüente de determinadas palavras na língua em detrimento de outras que raramente ou nunca co-ocorrem, embora tenham o mesmo conteúdo semântico. E mais a aquisição de um “significado” por parte dessas expressões consagradas, que pode ser desde próximo até totalmente distinto da “soma dos significados das palavras” que as compõem.

Não foi possível reconhecer no *output* que os corpora produziram – nem na natureza das dificuldades de tradução ligadas a essas expressões – qualquer tipo de diferenciação que justificasse a distinção em pauta. Nossa busca *corpus driven* gerou indistintamente soluções com colocações verbais ou com *phrasal verbs*, expressões idiomáticas e expressões fixas (vide Apêndice 1).

Isso não significa que essa distinção não seja justificada, e útil em certas aplicações. Porém ela não se origina, ao que tudo indica, em algo intrínseco à língua. *It does not carve language at a joint.*

7. Considerações Finais

Assim, os corpora emergem do presente trabalho como uma versátil e multifacetada ferramenta para a pesquisa de colocações verbais, especialmente para o tradutor, tolhido pela rigidez e escassez da lexicografia a respeito do assunto.

As diversas abordagens que adotamos permitiram a execução de uma série de tarefas: a busca e localização de colocações no corpus paralelo obtendo assim uma ou várias opções de tradução; a confirmação de que se está de fato diante de uma colocação verbal através da constatação de seu uso, freqüência e “identidade semântica” nos corpora; a confirmação da acepção corrente de uma colocação; a constatação de uma segunda acepção até então desconhecida; a busca a partir de indicações dos *collocate samplers*, revelando colocações insuspeitadas, incluindo palavras que nos eram até aqui desconhecidas; e um caso de “índice

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

sintático”, a preposição que se segue à colocação *have the heart* orientando o tradutor na definição da acepção exata.

Ademais, a busca *corpus driven* revelou a possibilidade de construir-se um “dicionário vivo e permanentemente atualizável”, que inclua todas as colocações de um determinado verbo de alta frequência na língua constantes no Compara. Até mesmo as de um verbo auxiliar, como *have*. E isso graças ao fato de que a sintaxe de busca do Compara demonstrou um surpreendente poder de expressão semântica, que permitiu que se restringisse as soluções a um *have* bastante “despido de sua dimensão gramatical”. O poder expressivo dessa sintaxe de busca merece ser mais bem estudado pelo tradutor. Ficou-nos mesmo a impressão de se estar diante de um potencial ainda não devidamente explorado.

Enfim, o uso de corpora na pesquisa das colocações verbais com vistas à tradução tornou viável e fascinante o que talvez sem eles fosse quase inviável ou extremamente árduo.

Mas não só para a execução desse tipo de tarefas serviram os corpora. Foi possível surpreender neles – ou surpreender-se com – situações que instanciam aspectos de interesse da epistemologia, da filosofia da linguagem, da filosofia da ciência. E, mais ainda, usá-los como uma espécie de “campo de prova” da relevância de conceitos empregados no estudo da língua. Como ferramentas através das quais conceitos que são intrínsecos à natureza da linguagem afloram, “*carving language at its joints*”, revelando a forma como a língua se articula.

To carve tem duas acepções. Por um lado, a de “esculpir, cinzelar, gravar, entalhar”. Por outro, a de “trinchar”: “cortar com certa elegância, em especial a carne que se servirá à mesa”. É essa a acepção que aqui se aplica. Quem “*carve at the joints*”, “trincha nas juntas”. Mas quem trincha – para servir à mesa – me informa que **sempre** se trincha nas juntas (quando há juntas). Portanto, a tradução da expressão que emprestamos de Hempel é, simplesmente, “trinchar a língua”.

E assim corpora se revelam “ferramentas de trinchar a língua”. Trinchar para traduzir. Trinchar para destrinchar.

Agradecimentos

A Diana Santos e ao pessoal do Compara, pelo suporte crucial no momento exato.

A Beatriz Souza, Marcia Fiker e Stela Foley, pela acolhida carinhosa e pelas discussões tão profícuas.

A Stella Tagnin, pelo entusiasmo e pelo apoio.

Bibliografia

- ALSTON, W. P. "Quine on Meaning", in Hahn, L.P. e Schilipp, P. A. eds., *The Philosophy of W.V. Quine*. La Salle, Open Court. 1988.
- BAKER, M. *In Other Words*. New York, Routledge. 1992.
- GENTZLER, E. *Contemporary Translation Theories*. New York, Routledge. 1993.
- HEMPEL, C. G. "Fundamentals of Taxonomy" in, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York, The Free Press. 1968.
- HUDDLESTON, R. D. & PULLUM, G. K., eds., *The Cambridge Grammar of the English Language*. Cambridge, Cambridge University Press. 2002.
- LONARDONI, M. "Aconteceu virou manchete – Um estudo dos vetores de manchetes jornalísticas", in VASCONCELOS, S. I. *Os discursos jornalísticos: reportagem, manchete, classificados & artigo*, Itajaí, Editora da Univali / Maringá, Eudem, 1999.
- MANSER, M. H., ed., *Good Word Guide*. Londres, Bloomsbury. 1990.
- QUINE, W. V. "Translation and Meaning", segundo capítulo de *Word and Object*. Cambridge, The MIT Press. 1960.
- SWAN, M. *Practical English Use*. Oxford, Oxford University Press. 2002.
- TAGNIN, Stella O. *Expressões Idiomáticas e Convencionais*. São Paulo, Ed. Ática. 1989.

Dicionários:

- DICTIONARY OF ENGLISH PHRASAL VERBS AND THEIR IDIOMS. London, Collins. 1990.
- NOVO DICIONÁRIO DA LÍNGUA PORTUGUESA. Coord, Aurélio Buarque de Holanda Ferreira, Rio de Janeiro, 1ª edição, Editora Nova Fronteira, s/d.
- OXFORD COLLOCATIONS DICTIONARY FOR STUDENTS OF ENGLISH. Oxford, Oxford University Press. 2002.
- THE CONCISE OXFORD DICTIONARY. Oxford, Clarendon Press. 1990
- APÊNDICE 1
- COLOCAÇÕES "NÃO VERBAIS" DE "HAVE" E "GO" OBTIDAS NO Comparação (CONSIDERADAS DE INTERESSE PARA O TRADUTOR)

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

COLOCAÇÕES DE “HAVE”

EBDL1T1 1(1520):	<i>She had it off with young Ritchie at the New Year's Eve.»</i>	Foi para a cama com o Ritchie mais novo na noite de fim de ano.
EBDL1T2 2(1520):	She had it off with young Ritchie at the New Year's Eve.»	Ela transou com o garoto Ritchie na festa de Ano Novo.
EBDL1T1 1(332):	<i>Doesn't waste time poking around in your unconscious, asking you about potty training, or whether you saw your parents having it off together, that sort of thing.»</i>	Não perde tempo a vasculhar o inconsciente, a fazer perguntas sobre a indiferença com que fomos criados, ou se alguma vez vimos os nossos pais na cama e esse tipo de coisas.»
EBDL1T2 2(2092):	<i>He has better grounds for complaint than me (should that be «I»?) - indeed, he could probably have me for assault.</i>	Ele tinha mais razões para reclamar do que eu (ou deveria dizer «para reclamar de mim»?) – na verdade, poderia me mandar prender por agressão à sua pessoa.
EBDL1T1(1764):	<i>«To be frank, you and Jake have us over a barrel on this one.»</i>	– Para falar com toda a franqueza, estamos à mercê de ti e do Jake e nesta questão.
EBDL1T2(1764):	<i>«To be frank, you and Jake have us over a barrel on this one.»</i>	– Para ser honesto, você e Jake nos têm pelo dedo mindinho nessa aqui.
EBDL1T2 2(146):	<i>«He's only having you on,» said the nurse.</i>	Ele está tirando sarro de você – disse a enfermeira.

COLOCAÇÕES DE “GO”

EBDL1T1 1(748):	No go .	Nem pensar .
EBIT1 (222):	<i>He was the only one of the Merediths to go in for cattle, too.</i>	Também era o único dos Merediths que criava gado .
ESNG1 (577):	<i>The subdued monosyllable was pronounced with such certainty; the habit of each other had made them even less demonstrative than they had been at the beginning of their marriage, but he was moved to go over to her.</i>	O monossílabo em voz baixa foi pronunciado com tal certeza; o hábito de cada um fizera-os menos afeitos a demonstrações que no começo de seu casamento, mas ele sentiu vontade de abraçá-la .
EBDL1T1 1(600):	<i>If she were to go with anyone else, something new in her behaviour, some unfamiliar adjustment of her limbs, some variation in her caresses, would tell me, I'm certain.</i>	Se ela andasse com alguém , tenho a certeza de que haveria algo de diferente no seu comportamento, uma inovação qualquer na sua forma de ajustar os membros, uma variação nas suas carícias, que me levaria a descobrir.
EBDL1T2 2(600):	<i>If she were to go with anyone else, something new in her behaviour, some unfamiliar adjustment of her limbs, some variation in her caresses, would tell me, I'm certain.</i>	Se ela tivesse dado para outro , alguma coisa nova no seu comportamento, alguma disposição nova de seus braços ou pernas, alguma variação no seu jeito de fazer carinho me diriam, disso estou certo.

APÊNDICE 2

Busca específica de colocações de “*have*” e “*go*” nos corpora: alguns exemplos adicionais

Adicionalmente aos exemplos constantes do corpo do artigo, apresentamos abaixo algumas das ocorrências obtidas nas buscas específicas por colocações nos corpora. Alguns dos casos, inclusive, não correspondem a colocações verbais, mas são apresentados por ilustrarem alguns pontos ressaltados no texto principal, ou por seu interesse para o trabalho de tradução. Cada grupo de exemplos é seguido de um breve comentário.

A) Exemplos 1, 2, 3, e 4

Exemplo 1

<http://www.patternlanguage.com/apl/aplsample/apl76/apl76.htm>

*Many small households, not large enough to have a full fledged nursery, not rich enough to have a nanny, find themselves swamped by the children. The children naturally want to be where the adults are; their parents don't **have the heart**, or the energy, to keep them out of special areas; so finally the whole house has the character of a children's room - children's clothes, drawings, boots and shoes, tricycles, toy trucks, and disarray*

Exemplo 2

http://1stholistic.com/Reading/liv_inspiration-the-heart-of-the-beginner.htm

*The Bible's very clear about it. While God does want us to grow in knowledge (Hosea 4:6), His desire is that we constantly **have the heart** of the Beginner. The heart of one who loves Him with all their heart, mind, strength and soul (Lk. 10:27). Oblivious to human praise. "The eyes of the Lord roam over the whole earth, to encourage those who are devoted to him wholeheartedly." (2 Chr. 16:9)*

Exemplo 3

<http://bnewheart.homestead.com/>

*Do you **have the heart** to be an organ donor?*

As of July 30, 2002 in the US there is a total of 80,076 persons waiting for organs.

Broken down as follows: 52,686 waiting for a kidney, 1,313 waiting for a pancreas,...

Exemplo 4

http://www.charlotte.com/mld/charlotte/sports/columnists/ron_green_sr/3709833.htm

Posted on Mon, Jul. 22, 2002 (Charlotte Observer)

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua.*

COMMENTARY

French **have the heart**, not the execution, for dares **RON GREEN SR.**
Ah, those Frenchmen. They are so, how you say, so ... French.

When Jean Van de Velde came to the last hole at Carnoustie in 1999, he needed only to make a double-bogey to become the first Frenchman since 1907 to win the British Open. Logic dictated he hit something like a 5-iron off the tee, then lay up short of the water that fronts the green with a wedge, hit another wedge onto the green and leave himself three putts to win

O exemplo 1 é claramente de um *collocate sense*. O exemplo 2 talvez não seja totalmente *plain*: “*to have the heart of the Beginner*” parece requerer algo de coragem (“*strength*”), mas a tradução é, certamente, “ter o coração...”. No exemplo 3, ao qual nos referimos de passagem no corpo do trabalho, há a alusão entre os *senses*, irrecuperável na tradução. O exemplo 4 é bem interessante: *you have one thing required “for dares” (“the heart”), but you miss the other (“the execution” = a técnica, o talento)*. Portanto, a estrutura sintática da frase puxa para o *plain* (ter “uma coisa” mas não ter “outra”), mas “*heart*” precisa ser traduzido como “*coragem*”. Um caso, diríamos, em que o “*in-between*” se situa não só no nível semântico, mas também “entre os níveis” sintático e semântico. Vide no corpo do artigo (seção 5) nossas observações sobre as diferentes preposições – *to, for* ou *of* – que se seguem a esta colocação servirem como um “índice da acepção”.

B) Exemplos 5, 6, 7 e 8

Exemplo 5

<http://www.fnal.gov/pub/inquiring/questions/danny.html>

Questions About Physics

How does an electron on a 2p orbital cross the node of nucleus?

To Fermilab,

*I have a question that needs answering for my chemistry class. In an atom, how does an electron on a 2p (bell shaped) orbital cross the node of the nucleus (a region of zero probability)? Does it **have something to do with** the wave properties of an electron? I cannot find the answer on the internet or in my chemistry book.*

Thank you very much.

Danny

Exemplo 6

<http://www.med.umich.edu/1libr/subabuse/tobacc12.htm>

Tobacco

Cigarette Substitutes - What to Take Along

Just as you pack for any trip, you can also pack for your journey to freedom from smoking! Remember all the reasons you used to reach for cigarettes? These will be the reasons you reach for your cigarette substitutes.

Crop, 10, 2005

To **have something to do with** your hands.
To have something to do while waiting in traffic jams or movie lines or
for slow service at a restaurant.

As a “prop” to relax.

As a reward.

Exemplo 7

http://www.gerismith.com/music/music_chain_01.html

Something to Do with Laurie

Laurie and I drove into the city last night

Mostly we talked about the loves in our lives

And I don't know why but I have something to do with Laurie

We have shared the same secrets, we have loved the same men

Our paths keep crossing again and again

I don't understand why I have something to do with Laurie

This chain wraps around

And we can't break these links as hard as we try

This train knows where we're bound

So I guess we should JUST SIT BACK AND ENJOY THE RIDE

Laurie and I cannot seem to forget

Our dances with partners that don't know the steps

And I don't know what yet, but I HAVE SOMETHING TO LEARN from Laurie

Cause I woke up this morning and everything had changed

My vantage points had been rearranged

*And I know it sounds strange, but it **had***

***Something to do with** Laurie*

Exemplo 8

<http://www.qwerty.com/hkr.htm>

*To me, the most challenging thing I can think of is coming up with software that is really interesting and helpful. And that's a lot more difficult than just breaking into a system and causing a mess. Although, to be completely honest, I did **have something to do with** a long-ago Christmas day when all the computers in a very large international network suddenly started playing **Jingle Bells** in unison on their printers. But that's another story...*

ABDULKADER, Inacio. *Carving Language at its Joints: Os Corpora como Ferramentas de (Des)Trinchar a Língua*.

“*Have something to do with*” não é colocação verbal, porém optamos por fazer constar estes exemplos neste apêndice já que neste caso, é claro, também existe um *plain* e um *collocate sense* da expressão, que muitas vezes é um problema para o tradutor em termos da decisão sobre qual dos dois *senses* ocorre num caso específico. O exemplo 5 é claramente *collocate* (“ter algo a ver com”, “ter alguma relação”), o texto técnico usa a expressão para indagar sobre a possibilidade de uma relação causal. Igualmente o exemplo 8 que confessa ter havido “alguma” responsabilidade, com uma ironia que indica que a responsabilidade foi “total”. A ocorrência no exemplo 6 é claramente *plain* (“ter algo para fazer com”), mas permite-nos inferir que “*smoking has something to do with* (*collocate*) *having something to do with* (*plain*) *one’s hands*”. No caso do exemplo 7, como traduzir o título da música? A ocorrência da expressão na última estrofe é claramente *collocate* dado o “*it*” que a precede. Mas e as duas ocorrências da primeira estrofe como ficam? Um empático “ter a ver” ou algo mais físico “a fazer”? Dado o “sentido geral” da letra, intencionalmente (?) dúbio (há ou não uma alusão? vide as duas frases em caixa negrito), ficam essas duas ocorrências, como o título, “indeterminadas”. Cada leitura é uma leitura, e aqui prevalecerá a leitura do tradutor.