
CORPORA
NA TERMINOLOGIA

Corpora na Terminologia

© HUB Editorial 2013

HUB EDITORIAL

Rua das Palmeiras, 214

01226-010 São Paulo SP

Brasil

Tel.: 55 -11 - 3661 -1150

E-mail: contato@hubeditorial.com.br

Website: www.hubeditorial.com.br

Todos os direitos reservados. Nenhuma parte deste manual pode ser reproduzida ou usada sob qualquer forma ou meio, sem consentimento por escrito por parte dos Editores.

Coordenação Editorial

Susanna Florissi

Assistentes Editoriais

Ana Carolina Montecinos

Karina Kempler

Revisão do Português

Simone Zaccarias

Projeto Gráfico/Editoração

Cia. de Desenho

ISBN 978-85-8076-134-4

13-099525

Corpora na Terminologia / organização Stella Tognin e Cleci Bevilacqua . --
São Paulo : Hub Editorial, 2013.

Vários autores

1. Análise linguística 2. Corpora (Linguística) 3. Linguagem e línguas 4. Linguística computacional
5. Linguística - Metodologia 6. Terminologia I. Tognin, Stella. II. Bevilacqua, Cleci.

1. Linguística de corpus : Terminologia : Linguística aplicada 410,285

CDD-410.285

A LINGUÍSTICA DE CORPUS COMO METODOLOGIA
PARA A COMPILAÇÃO DE UM GLOSSÁRIO DE
TERMOS DA CULINÁRIA TÍPICA BRASILEIRA

Rozane Rodrigues Rebechi (USP)¹

ABSTRACT: This chapter presents the methodology used in building a Portuguese-English glossary of Brazilian cooking terms for translators and food writers in general. The interest in this area arose from the fact that such a work does not exist for this pair of languages. As a result, the terminology in the field is not standardized in translations and writings on this topic. In order to reach our goal, we compiled a comparable corpus of Brazilian cookbooks originally written in Portuguese and in English. The Portuguese subcorpus has 241,720 tokens, and the subcorpus in English, 427,193. With the help of the software *Wordsmith Tools 6.0*, Brazilian terms were retrieved from the Portuguese subcorpus, and English equivalents, definitions, examples, collocates, comments and other information were retrieved from the English subcorpus. Results show that authentic LSP texts are useful for building a bilingual terminological work, since terms, phraseologies and definitions can be retrieved from their real contexts.

KEYWORDS: Brazilian Cooking; LSP; Glossary; Corpus Linguistics; Terminology

1) Introdução

Apesar de já ter sido considerada como arte menor devido a sua materialidade (cf. CAPARTTI e MONTANARI, 1999), a culinária assume um papel cultural central para alguns estudiosos. Lévi-Strauss (2007) e Montanari (2009), por exemplo, equiparam-na à linguagem como expressão cultural. O interesse pelo tema também pode ser observado em estudos acadêmicos realizados nas duas últimas décadas, como os de Rebechi (2010a, 2010b e 2012), Moraes (2010), Camargo e Valdíório (2009), Costa (2006), Taghin e Teixeira (2004), Teixeira (2004 e 2008) e Corrêa et al (2003), entre outros.

¹ Processo FAPESP nº 2011/19609-0.

Em relação à culinária tipicamente brasileira, o crescente interesse, tanto por parte dos brasileiros quanto dos estrangeiros, parece estar diretamente ligado à visibilidade que vem conquistando dentro e fora do país, impulsionada por *chefs* como Alex Atala e Helena Rizzo², entre outros nomes que têm colocado o Brasil na rota da alta gastronomia mundial. Em ritmo acelerado observa-se, também, a publicação de livros de receitas brasileiras em língua inglesa. Em 2010, a livraria virtual *Amazon* comercializava treze títulos dedicados ao tema. Atualmente, disponibiliza pelo menos 24³.

Semelhante intensidade, contudo, não é observada em relação à publicação de materiais terminográficos bilíngues na área. Quanto à culinária geral, poucas são as obras de referência no par de línguas português-ínglês dedicadas ao tema. Entre elas, vale destacar o *Glossário de Gastronomia Português-Ínglês/Ínglês-Português* (KLIEB, 2006), o *Dicionário Gastronômico português-espanhol/ínglês/alemão/francês/italiano* (CARLI e KLOTZ, 2007) e o *Vocabulário para Culinária inglês/português* (TAGNIN e TEIXEIRA, 2008). Em relação à culinária tipicamente brasileira, não existe obra terminográfica publicada que forneça equivalentes e/ou definições, em língua inglesa, para termos próprios dessa subárea. Essa lacuna tem resultado em falta de padronização, definições incorretas, erros ortográficos, descaracterização de aspectos culturais etc. (cf. REBECHI, 2010a). Problemas encontrados tanto em material traduzido para o inglês quanto em material escrito originalmente nesse idioma.

Como o trabalho do terminólogo deve ser orientado pelas necessidades de expressão dos usuários (DUBUC, 1999), a fim de prestar uma contribuição para tradutores e redatores que tratam da culinária típica brasileira em inglês, estamos compilando um glossário português-ínglês da área, a partir de pressupostos da Terminologia e da Linguística de *Corpus*. Tal obra tem como objetivo auxiliar esses profissionais a produzir textos fluentes em língua inglesa, quer sejam traduções ou produções nessa língua. Este capítulo descreve a metodologia utilizada nessa investigação.

² Chefs-proprietários dos restaurantes D.O.M. e Maní, respectivamente, que estão entre os 50 melhores do mundo, segundo ranking divulgado pela revista *Restaurant* em 2012.

³ Segundo consulta realizada no site *www.amazon.com* em 20 julho de 2012.

2) Receitas culinárias como fonte de extração terminológica

Em geral, mas não exclusivamente, termos são encontrados em situações comunicativas especializadas (MACHEL, 2001). Talvez os escassos estudos em terminologia culinária sejam resultado de uma justificativa provavelmente familiar àqueles que pesquisam temas ditos 'cotidianos': o discurso especializado seria restrito a um grupo de profissionais de determinada área, ao passo que a culinária seria tema do dia-a-dia e, portanto, acessível a qualquer falante. Se assim fosse, receitas não seriam consideradas como textos técnicos e não poderiam servir como fonte de extração terminológica. Dubuc (1999), contudo, apresenta uma visão bastante abrangente sobre línguas de especialidade. Para o pesquisador, a palavra 'técnico' abrange quase todas as atividades humanas, entre elas, as artes, as ciências, as profissões, as atividades de lazer etc.

Devido a sua importância no cotidiano dos indivíduos, a terminologia alimentar possui grande penetrabilidade na linguagem geral e, apesar da não uniformidade, dadas as variadas condições de uso em que é produzida, não deixa de ser 'comunicação especializada', pois possui características específicas (BACCIN, 2003).

A Tabela 1 apresenta uma lista das palavras mais frequentes, em ordem decrescente, extraídas de um *corpus* de receitas culinárias em português, que possui mais de um milhão e meio de palavras⁴.

⁴ Esse *corpus* foi compilado por Teixeira (2008) e, nesta pesquisa, foi utilizado também como *corpus* de referência em português.

N	Word	Freq.	%	Texts	%
1	DE	131.296	8,37	6.931	99,97
2	#	117.171	7,47	6.905	99,6
3	E	68.012	4,34	6.696	96,58
4	O	47.383	3,02	6.319	91,14
5	A	43.595	2,78	6.390	92,17
6	COM	30.184	1,92	6.278	90,55
7	EM	28.381	1,81	6.264	90,35
8	SOPA	15.791	1,01	5.269	76
9	POR	15.529	0,99	5.154	74,34
10	UMA	13.979	0,89	4.837	69,77
11	SAL	13.279	0,85	4.564	65,83
12	AS	12.100	0,77	4.240	61,16
13	ATÉ	11.944	0,76	4.422	63,78
14	CHÁ	11.740	0,75	4.039	58,26
15	OS	11.151	0,71	4.326	62,4
16	PARA	10.912	0,7	4.320	62,31
17	MINUTOS	10.601	0,68	4.322	62,34
18	PREPARO	10.508	0,67	6.473	93,37
19	COLHERES	10.192	0,65	4.346	62,69
20	DO	10.089	0,64	4.190	60,44
21	INGREDIENTES	9.811	0,63	6.882	99,26
22	UM	9.645	0,61	3.923	56,58
23	XÍCARA	9.461	0,6	3.437	49,57
24	ÁGUA	9.366	0,6	3.924	56,6
25	COLHER	9.360	0,6	3.902	56,28
26	AO	8.614	0,55	4.174	60,2
27	COLOQUE	8.284	0,53	3.936	56,77
28	FOGO	8.242	0,53	3.295	47,53
29	OU	7.893	0,5	3.545	51,13
30	LEITE	7.833	0,5	2.505	36,13

Tabela 1

Primeiras 30 palavras mais frequentes de um *corpus* de receitas culinárias em geral.

Apesar de as palavras mais frequentes do *corpus* (sopa⁵, sal⁵, colher(es)⁵, chá⁵, minutos⁵, preparo⁵ etc.) fazerem parte, também, do léxico comum dos falantes de português, elas, em conjunto, nos remetem imediatamente à linguagem das receitas culinárias.

Assim como em outros campos do saber, na culinária também são encontradas palavras da língua comum que passaram por um processo de “terminologização” (cf. PINATTO e KRIEGER, 2004, p. 79). Como exemplo, podemos citar ‘xícara’, palavra que ocupa a 23^a posição em frequência no nosso *corpus* de estudo em português. A primeira acepção da palavra no *Dicionário Aurélio* (FERREIRA, 1986) é: “pequena vasilha com asa para servir em especial bebidas quentes [...]”. Já em receitas culinárias, ‘xícara’ representa uma quantidade equivalente a 250 ml, o que vem confirmar a conclusão de Dubuc (1999, p. 57): “De acordo com as necessidades, a língua de especialidade amplia, restringe ou modifica o significado das palavras da língua geral”⁵.

3) Metodologia

Pesquisas terminológicas sempre foram baseadas em *corpus*, ou seja, em um conjunto de textos da área em questão. No entanto, o trabalho que, até duas ou três décadas atrás, era feito manualmente, passou a contar com a ajuda fundamental de ferramentas computacionais (cf. ALMEIDA e CORREIA, 2008). Portanto, atualmente muitas pesquisas terminológicas aliam Terminologia e Linguística de *Corpus* (ver, entre outros, PICHT, 2004; ALMEIDA e CORREIA, 2008; ORENHA-OTTALIANO, 2008; ZILIO, 2008).

Um *corpus* pode servir para extração terminológica desde que tenha sido compilado criteriosamente e, portanto, seja representativo tendo em vista os objetivos da pesquisa. Tomando como base os aspectos metodológicos envolvidos na compilação de um *corpus* para a finalidade terminográfica apontados por Almeida e Correia (2008), as seguintes questões foram consideradas para a compilação do *corpus* de estudo: delimitação do domínio; seleção das fontes; compilação dos textos; conversão de formatos; nomeação dos arquivos; e anotação dos textos.

⁵ No original: “Segun sus necesidades, la lengua de especialidad amplia, restringe o modifica el significado de las palabras de la lengua general”.

3.1) O *corpus* de estudo

Para servir como fonte de extração de dados para a construção do glossário de termos da culinária brasileira, foi compilado um *corpus* comparável⁶ bilingue português-inglês a partir de livros de receitas escritos originalmente nos dois idiomas⁷. Todo o conteúdo textual desse material foi digitalizado e salvo em formato *.txt* para ser passível de análise semiautomática com ajuda da ferramenta *WordSmith Tools 6.0* (doravante, *WSTR*) (SCOTT, 2012). Uma vez que os livros utilizados não se limitam a descrever receitas, mas também incluem comentários, curiosidades, informações históricas etc., que podem ser úteis para a construção de definições, comentários e exemplos dos verbetes, esse conteúdo, denominado 'informativo', também foi preservado e salvo como um *subcorpus* que pode ser acessado quando necessário.

Em relação ao número de palavras, o *corpus* de estudo está assim distribuído:

	Subcorpus em português	Subcorpus em inglês
Receitas	166.296	279.057
Textos informativos	75.424	148.136
Nº total de palavras	241.720	427.193

Tabella 2
Número de palavras do *corpus* de estudo.

Apesar de os dois *subcorpora* não estarem balanceados quanto ao número de palavras, essa diferença de extensão não parece prejudicar a pesquisa, uma vez que o *subcorpus* em português serve, principalmente, como fonte de extração de termos. Possíveis equivalentes, definições, exemplos e outras informações relevantes serão extraídos do *subcorpus* em inglês, que deve ser tão extenso quanto possível (cf. AUBERT, 1996).

⁶ A definição de *corpus* comparável utilizada é a de textos originais nos idiomas envolvidos (cf. BOWKER e PEARSON, 2002).

⁷ A lista completa dos livros que compõem o *corpus* encontra-se no Anexo.

3.2) Levantamento dos termos

A fim de salientar o que é característico de um dado *corpus*, ou seja, suas palavras-chave, deve-se comparar a lista de palavras desse *corpus* com a de um *corpus* de referência. Neste estudo, observamos que a comparação do *corpus* de estudo com um *corpus* de língua geral, por exemplo, poderia não salientar o que é característico da culinária típica brasileira, mas sim o que é próprio da culinária como um todo. Portanto, a lista de palavras do *subcorpus* de estudo em português foi comparada à de um *corpus* de receitas em geral, no mesmo idioma. A partir das palavras-chave resultantes dessa comparação, estão sendo extraídos os candidatos a termos do glossário.

A Tabela 3 mostra as primeiras dez palavras-chave do *subcorpus* em português, apresentadas em ordem decrescente de chavicidade (*keyness*)⁸.

N	Key word	Freq.	%	RC.	Freq.	RC.	%	Keyness	P
1	CHEIRO-VERDE	308	0,18	143	0,01	849,18	0,000000000000		
2	MANDIOCA	409	0,24	387	0,03	81,741	0,000000000000		
3	COCCO	528	0,31	1.055	0,07	588,31	0,000000000000		
4	DÁ	220	0,13	134		546,43	0,000000000000		
5	KG	678	0,4	1.837	0,13	520,61	0,000000000000		
6	AZEIFE-DE-DENDÊ	141	0,08	48		428,45	0,000000000000		
7	NUMA	878	0,52	3.472	0,25	340,83	0,000000000000		
8	CAMARÃO	297	0,18	652	0,05	298,18	0,000000000000		
9	PIRÃO	120	0,07	81		284,92	0,000000000000		
10	COENTRO	275	0,16	593	0,04	281,76	0,000000000000		

Tabella 3
Primeiras 10 palavras-chave do *subcorpus* em português.

⁸ Para calcular o índice de chavicidade de uma palavra, o programa compara estatisticamente os seguintes índices: (i) a frequência da palavra no *corpus* de estudo, (ii) o total de palavras (*tokens*) do *corpus* de estudo, (iii) a frequência da palavra no *corpus* de referência, e (iv) o total de palavras do *corpus* de referência (SCOTT, 2012). Quanto mais estatisticamente diferente no *corpus* de estudo, maior será a 'chavicidade' da palavra.

Por meio da *Keyword List* apresentada pelo programa WST, é possível identificar a frequência com que a palavra-chave aparece no *corpus* de estudo (coluna *Freq.*), o percentual de vezes em que essa palavra ocorre em comparação às outras palavras do *corpus* (coluna %), o número de vezes em que a palavra ocorre no *corpus* de referência (coluna *RC. Freq.*) e a que percentual esse número corresponde (*RC. %*), a chavidade da palavra (coluna *Keywords*) e o valor de p^2 (coluna *P*). Em relação ao nosso *subcorpus* em português, observamos que as primeiras palavras-chave são, de fato, utilizadas com frequência significativamente maior do que no *corpus* de referência, uma vez que o valor de p de todas é 'zero', demonstrando que a probabilidade de a palavra ter sido considerada chave ao acaso é nula (cf. ROCHA, 2007). No entanto, isso não significa que todas essas palavras constituam candidatos a termos da culinária típica brasileira.

Podem-se observar, mesmo entre as primeiras palavras-chave, algumas que são utilizadas em receitas em geral. É o caso de 'dá', 'kg', 'numa', 'tire' etc. A análise da lista de palavras do *corpus* de referência revelou uma preferência pelas variantes 'rendimento', 'g', 'em uma', 'reire' etc., o que explica o porquê de palavras da culinária geral aparecerem entre as palavras-chave do *corpus* de estudo. Vale ressaltar também, que o *subcorpus* de estudo em português e o *corpus* de referência nesse idioma foram compilados a partir de fontes diferentes: o primeiro provém de livros de receitas, enquanto o último é composto de receitas extraídas de sites da Internet. Isso pode justificar a variação de registro observada. De qualquer forma, nesta pesquisa nosso interesse recai sobre as palavras e combinações de palavras que se referem a ingredientes e pratos, excluindo-se, portanto, verbos, palavras gramaticais, palavras que se referem a pesos e medidas etc.

Para realizar o levantamento das palavras do *subcorpus* de estudo em português, a ferramenta foi ajustada para considerar palavras separadas por hífen como uma palavra apenas. Portanto, em relação às palavras de conteúdo, observam-se, entre as primeiras palavras-chave, unidades lexicais hífenizadas ('cheiro-verde', 'azeite-de-dendê' etc.) e simples ('mandioca', 'coco', 'canarão'), entre outras.

A compilação de um glossário português-inglês de culinária típica brasileira poderia simplesmente listar os termos-chave em português

9 A probabilidade de um evento ocorrer por acaso varia entre 0 e 1, sendo que 0 significa que não há chance de ocorrer por acaso e 1 significa a certeza de que ocorre (OAKES, 1998).

('cheiro-verde', 'mandioca', 'coco' etc.), que formariam as entradas do material terminográfico, e alinhá-los a seus 'equivalentes' em inglês (*parsley and green onion, manioc, coconut* etc.)¹⁰, o que já não seria uma tarefa fácil, uma vez que não há equivalentes padronizados para todos esses termos. Porém, apenas conhecer os equivalentes na língua de chegada não garante a produção de um texto fluente. Em geral, as palavras vêm acompanhadas de outras, formando 'colocações' (cf. BERBER SARDINHA, 2004; TAGNIN, 2013) e fraseologias.

As fraseologias especializadas, ou Unidades Fraseológicas Especializadas (UFE), definidas por Bevilacqua (2001, p. 106) como "unidades sintagmáticas que incluem um termo entre seus elementos, que possuem um determinado grau de fixação e uma frequência relevante em um conjunto de textos ou em um âmbito especializado", vêm chamando mais atenção, uma vez que tradutores e redatores precisam utilizar os termos de forma a produzir textos adequados na língua de chegada. Portanto, no glossário de termos da culinária típica brasileira pretendemos oferecer ao tradutor/redator exemplos do termo em seu contexto real de uso, assim como suas principais combinações e fraseologias.

A título de ilustração, tomemos como exemplo o termo com maior chavidade no nosso *corpus*, 'cheiro-verde'. Por meio da função *collocates* do utilitário *Concord* do WST, podemos visualizar seus principais colocados:

Collocates	freq.	total	freq.	total
1. cheiro-verde	176	200	0	0
2. cheiro-verde	144	171	11	22
3. cheiro-verde	129	212	11	22
4. cheiro-verde	126	148	11	22
5. cheiro-verde	125	112	41	18
6. cheiro-verde	120	169	14	22
7. cheiro-verde	110	169	15	24
8. cheiro-verde	110	112	11	22
9. cheiro-verde	109	112	11	22
10. cheiro-verde	109	112	11	22
11. cheiro-verde	109	112	11	22
12. cheiro-verde	109	112	11	22
13. cheiro-verde	109	112	11	22
14. cheiro-verde	109	112	11	22
15. cheiro-verde	109	112	11	22
16. cheiro-verde	109	112	11	22
17. cheiro-verde	109	112	11	22
18. cheiro-verde	109	112	11	22
19. cheiro-verde	109	112	11	22
20. cheiro-verde	109	112	11	22
21. cheiro-verde	109	112	11	22
22. cheiro-verde	109	112	11	22
23. cheiro-verde	109	112	11	22
24. cheiro-verde	109	112	11	22
25. cheiro-verde	109	112	11	22
26. cheiro-verde	109	112	11	22
27. cheiro-verde	109	112	11	22
28. cheiro-verde	109	112	11	22
29. cheiro-verde	109	112	11	22
30. cheiro-verde	109	112	11	22

Figura 1
Principais colocados do termo 'cheiro-verde'.

10 Neste momento da pesquisa, os equivalentes sugeridos servem unicamente como exemplos. Não estamos considerando variação geográfica, sinonímia e outras questões que seriam tratadas nos verbetes do glossário.

Observamos que, entre as 308 ocorrências de 'cheiro-verde' (1ª linha), em 84 (coluna R1) o termo possui, como primeiro colocado à direita, a palavra 'picado', participio que assume a função de adjunto adnominal, e em 60 (coluna L2), como segundo colocado à esquerda, a palavra 'maço', indicadora de quantidade. Agora vejamos o que a função *Clusters*, do WST, revela como combinações mais frequentes com o termo 'cheiro-verde':

The screenshot shows the WST interface with the search term 'cheiro-verde'. The main window displays a list of clusters with columns for rank (N), cluster name, frequency (F), and position (S). The top cluster is 'cheiro-verde picado' with a frequency of 111. Other clusters include 'maço de cheiro-verde', 'cheiro-verde e', and 'maço de cheiro-verde'. The interface includes menu options like 'File', 'Edit', 'View', 'Compute', 'Settings', 'Windows', and 'Help'.

Figura 2
Combinações frequentes com o termo 'cheiro-verde'.

Esse levantamento confirma que as UFEs 'maço de cheiro-verde' (60 ocorrências) e 'cheiro-verde picado' (84 ocorrências) são frequentes no *corpus* e, portanto, devem ser resgatadas nas receitas em inglês.

A segunda palavra-chave do sub*corpus* é 'mandioca', ingrediente que qualquer conhecedor de culinária brasileira concordaria que deve receber entrada no nosso glossário por se tratar de importante elemento da nossa dieta. Porém, não é só a unidade simples 'mandioca' que se caracteriza como termo. Aliás, retomando Dubuc (1999), em Terminologia, as formas sintagmáticas são dominantes, uma vez que a unidade terminológica está ligada à necessidade de expressar as realidades próprias de uma área em situação concreta de funcionamento. Vejamos as combinações mais frequentes com o termo 'mandioca':

The screenshot shows the WST interface with the search term 'mandioca'. The main window displays a list of clusters with columns for rank (N), cluster name, frequency (F), and position (S). The top cluster is 'farinha de mandioca' with a frequency of 357. Other clusters include 'farinha de mandioca e', 'mandioca e', and 'farinha de mandioca cruá'. The interface includes menu options like 'File', 'Edit', 'View', 'Compute', 'Settings', 'Windows', and 'Help'.

Figura 3:
Combinações mais recorrentes com a palavra de busca 'mandioca'.

Por meio do levantamento das combinações pela função *Clusters* do WST, observamos que, das 409 ocorrências de 'mandioca' (canto inferior esquerdo), 299 são de 'farinha de mandioca' (linha 3). A partir dessas 299 ocorrências, observam-se, ainda, em linhas que fogem da área de visualização da figura 3, sintagmas maiores, tais como 'farinha de mandioca torrada' (12 ocorrências), 'farinha de mandioca cruá' (12 ocorrências) e 'farinha de mandioca fina' (6 ocorrências). Considerando-se a extensão do sub*corpus* de estudo em português, podemos afirmar que essas unidades correspondem a fortes candidatos a termos do glossário, seja como verbetes ou subverbetes. No entanto, devemos lembrar que a ferramenta computacional nos mostra as combinações mais recorrentes, mas não tem capacidade cognitiva para determinar o que de fato é uma unidade terminológica. Somente a interface "ferramenta - homem" (PICHT, 2004) é capaz de identificar, por exemplo, que as combinações 'de mandioca' (linha 1), 'farinha de' (linha 2), 'de farinha' (linha 4) etc. não formam UFEs, mas 'farinha de mandioca torrada/fina/cruá', 'massa de mandioca' etc., sim.

3.3) Preenchimento do enunciado terminográfico

Conforme observado anteriormente, o glossário de termos da culinária brasileira aqui descrito é um produto terminográfico bilingue na direção português-inglês que tem, como principal público-alvo, tradutores e redatores que produzem, em inglês, textos que abordam a culinária tipicamente brasileira. Portanto, após a identificação dos termos (em português) que receberão entrada na obra¹¹, partimos para o preenchimento do enunciado terminográfico (em inglês) que, neste estudo, engloba a busca por equivalentes e/ou definições, informações enciclopédicas, exemplos e outras informações julgadas relevantes para esses profissionais.

Pelo fato de a culinária configurar-se como aspecto culturalmente marcado, é de se esperar que essa busca por equivalência vá muito além da mera identificação de 'títulos' em língua inglesa para conceitos em português, pois a equivalência completa entre códigos nem sempre é possível (cf. JAKOBSON, 1959). O andamento da pesquisa vem nos mostrando que, muito frequentemente, os conceitos próprios à culinária tipicamente brasileira são desconhecidos para a maioria dos falantes de língua inglesa e, conseqüentemente, não possuem equivalentes consagrados naquele idioma. Nesses casos de não equivalência direta, Aubert (1996) propõe que os principais traços conceituais do termo sejam resgatados por meio de definições, explicações ou descrições.

A fim de encontrar possíveis equivalentes para os termos que receberão entrada no glossário, o primeiro passo foi identificar, no *subcorpus* em português, receitas que geralmente utilizam o ingrediente em questão. Em seguida, pesquisamos, no *subcorpus* em inglês, como essas mesmas receitas abordaram o termo.

Vale ressaltar que, por uma limitação de espaço, neste estudo trataremos apenas dos dois termos com maior chavidade no *subcorpus* em português, 'cheiro-verde' e 'mandioca'.

Idealmente, para cada conceito haveria um único termo que o denominasse. Nesse caso, não haveria sinonímia ou homonímia em

Terminologia. Porém, é de conhecimento geral que, na prática, isso não acontece. Tanto na língua geral quanto na linguagem de especialidade, muitas vezes surgem diferentes termos para nomear um mesmo conceito e, em outras, um só termo nomeia diferentes conceitos. Essa falta de univocidade ficou bastante evidente durante a construção das duas entradas aqui descritas ('cheiro-verde' e 'mandioca'), conforme veremos adiante. Dubuc (1999) acredita que é dever do terminólogo tentar eliminar o excesso de sinonímia, a fim de preservar a eficácia da comunicação; quando isso não for possível, que estabeleça o verdadeiro alcance de cada sinônimo. Essa questão será contemplada no glossário que está sendo construído.

A fim de construir um glossário especializado, devem-se estabelecer os paradigmas mínimos (constantes em todos os verbetes) e possíveis da microestrutura (cf. BARROS, 2004). Durante a construção das fichas terminológicas dos verbetes aqui descritos, foram estabelecidos os seguintes elementos como mínimos na obra que está sendo construída:

- a. Entrada na língua de partida (LP);
- b. Classificação gramatical;
- c. Equivalente(s) na LC¹²;
- d. Definição na língua de chegada (LC);
- e. Exemplo(s) de uso na LC.

Em relação aos elementos possíveis, foram identificados:

- a. Nome científico;
- b. Principais combinações com o termo;
- c. Informações enciclopédicas;
- d. Remissivas.

¹¹ Por se tratar de pesquisa em andamento, ainda não foi determinada a quantidade de termos que serão contemplados na obra.

¹² Em caso de não equivalência direta, o termo será mantido em português.

Além desses, também foi observada a necessidade de acrescentar um elemento que chame a atenção do consultante para possíveis mal-entendidos. Observou-se, por exemplo, que no *subcorpus* em inglês foi utilizado *yucca* como equivalente de 'mandioca/alpim/macaxeira'. No entanto, com essa grafia o termo se refere a uma planta ornamental. Para o ingrediente, o correto seria *yuca*.

O *subcorpus* em inglês, em especial os textos informativos, tem-se mostrado bastante eficiente para a extração de definições e informações enciclopédicas. Nesse material, fica bastante evidente a relação entre o autor e o leitor, à qual Pearson (2004) se refere: escrevendo para um público que, em sua maioria, desconhece aspectos próprios à culinária brasileira, os autores desses livros procuram explicitar conceitos, facilitando, assim, a compreensão por parte dos leitores. Dessa forma, não raro encontramos explicações como:

Manioc: This starchy root, referred to in the States as "cassava" or "yuca," is called alpim in Rio and the northeastern states of Pernambuco and Bahia, macaxeira in the northeast (the Amazonian states), and mandioca in São Paulo and the southern states (Brazil: a Cook's Tour)

Tais informações podem ser utilizadas, no todo ou em partes, para a construção do enunciado definicional.

4) Modelos de verbetes

Em seguida, mostraremos a construção dos verbetes 'cheiro-verde' e 'mandioca', que vimos discutindo ao longo deste capítulo.

4.1) 'cheiro-verde'

A palavra com maior índice de chavricidade no *subcorpus* em português confirma as dificuldades envolvidas na compilação de um glossário bilíngue de termos da culinária típica brasileira. Em primeiro lugar, o termo não se refere a apenas um conceito por todo o território brasileiro. A definição apresentada em um dos dicionários especializados consultados – "denominação que é dada à combinação de salsa e cebolinha picadas e misturadas, utilizadas juntas no tempero de saladas, guisados, refogados etc." (ALGRANTI, 2004) – revela alguns problemas: em várias regiões do Brasil, em especial em estados do Norte, Nordeste e Centro-Oeste, 'cheiro-verde' refere-se à combinação de 'coentro e cebolinha', e não de 'salsa e cebolinha', como no Sul e no Sudeste. Além disso,

conforme mencionado anteriormente, essas ervas são, em geral, compradas em maços, e não previamente picadas.

A segunda dificuldade advém do fato de que, em inglês, não existe um termo apenas que englobe os dois ingredientes. Além disso, no *subcorpus* em língua inglesa, observou-se o uso de dois equivalentes diferentes para 'cebolinha': *scallion* e *green onion*. Isso vem comprovar que, se não há univocidade de terminologia dentro de uma mesma língua, o contato entre diferentes idiomas só faz aumentar a falta de padronização.

Visto tratar-se de uma obra terminográfica, optou-se por criar diferentes entradas para termos que possuam diferenças conceituais. A ordem em que esses termos vão aparecer na obra dependerá do número de ocorrências de cada um no *corpus* de estudo. A partir dessas e outras constatações, e, sempre que possível, utilizando o *corpus* como fonte de extração de dados, preenche-se a ficha terminológica de cada verbebe. Em geral, a construção das definições e das informações enciclopédicas está sendo feita a partir de diversos contextos retirados do *corpus*. Quando julgadas insuficientes, essas informações são complementadas por pesquisas feitas na Internet, em dicionários, glossários e outros materiais. Já os exemplos são extraídos do *corpus* e, portanto, fazem referência à fonte do mesmo. A figura 4 apresenta o verbebe 'cheiro-verde' (uma entrada e duas subentradas), construído por meio do software *IshwanLex*¹³.

cheiro-verde(noun) A combination of aromatic herbs used for flavoring stews, sauces, savory fillings etc. that is usually sold as a tied bunch similar to a bouquet garni, but mostly used finely chopped.
cheiro-verde(noun) parsley and green onion (s and se)//Ex.: *Heat the skimmed pan drippings over medium heat, add the chopped parsley and green onion, and sauté for 1 to 2 minutes.* [TCOB]//Clust:chopped (fresh) parsley and green onion; parsley and green onion, chopped; mixed parsley and green onion. See also **cebolinha**, **salsinha**.
cheiro-verde(noun) cilantro and green onion (N, NE and CW)//Ex.: *Return to the pot and add the olive oil, garlic, onion, coconut milk, green onions, cilantro, and salt, stirring until the mixture returns to a boil.* [BCJ]//Clust:green onions and cilantro, chopped. See also **cebolinha**, **coentro**.

Figura 4

Modelo de verbebe: 'cheiro-verde'.

¹³ Trata-se de um conjunto de aplicativos para construção e gerenciamento de materiais terminográficos e lexicográficos que pode ser adquirido pelo site: <http://ishwanlex.com/ishwanlex/>.

4.2) 'mandioca'

O preenchimento da ficha do termo 'mandioca' evidenciou ainda mais dificuldades. Assim como 'cheiro-verde', 'mandioca' também não possui apenas um conceito em português. Para se referir à raiz da família *Manihot esculenta*, que pode ser ingerida apenas cozida, são utilizados, pelo menos, três termos: 'mandioca', 'aipim' e 'macaxeira'. Na entrada de 'mandioca', em Algranti (2004), os três termos são tratados como sinônimos, e 'mandioca-mansa' e 'mandioca-brava' seriam seus hipônimos.

De fato, poderíamos afirmar que, segundo classificação de Barros (2004), 'mandioca', 'aipim' e 'macaxeira' seriam quase-sinônimos geográficos, uma vez que são utilizados em diferentes regiões do Brasil para se referir a um mesmo produto, também conhecido como 'mandioca-doce'. Porém, o termo 'mandioca' é utilizado na maior parte do país para se referir a outro tipo de raiz, a mandioca-brava, também da família *Manihot esculenta*, mas que possui um índice muito mais elevado de ácido cianídrico, fazendo com que essa raiz tenha de passar por processos que eliminem essa substância antes de ser transformada em diversos tipos de farinha e outros derivados. Portanto, o termo 'mandioca' no Nordeste, por exemplo, possui um conceito diferente daquele do Sudeste.

Além da falta de padronização do termo em língua portuguesa, observam-se, também, três diferentes equivalentes utilizados em língua inglesa: *manioc*, *cassava* e *yuca*. No entanto, também não é possível afirmar que sejam sinônimos, uma vez que não são intercambiáveis em todos os contextos. No *corpus* de estudo em inglês, *manioc*, *cassava* e *yuca*¹⁴ foram utilizados como sinônimos da raiz *in natura* e nas receitas que a utilizam simplesmente cozida, ou seja, como possíveis equivalentes de 'mandioca doce'. No entanto, o termo *yuca* não foi utilizado para se referir à raiz em forma de farinha, ou seja, não foram encontradas ocorrências de *yuca flour/meal*. Nesses casos, *manioc meal/flour* e *cassava meal* foram utilizados como sinônimos.

Considerando essas variações de uso, foi construída uma entrada e duas subentradas para o verbete 'mandioca':

¹⁴ Em quase todas as ocorrências do *corpus* em inglês o termo está erroneamente grafado *yuca*.



mandioca(noun) (*Manihot utilisima*) manioc, cassava, yuca A tropical plant with green leaves and starchy tuberous roots.//Basically two types are used: a non-poisonous type, which is known as *aipim* in Rio and the southern states, *macaxeira* in the North and Northeast and *mandioca* in São Paulo, and a poisonous type, known as *mandioca* throughout Brazil.

mandioca(noun) manioc, cassava, yuca Edible root which must be cooked before being fried or puréed.//Ex.: *Add the cooked manioc to the pan and shake a couple of times to coat the pieces with the butter and the onion.* [TBT]Ex.: *Peel cassava root, cut into slices and boil in water with 1 tablespoon salt until it is fork tender.* [Braz.Cook.//The non-poisonous tubers are simply cooked and fried or mashed, as they are used in the preparation of stews, such as *bobó*.//Clust:manioc root, (peeled); sweet manioc; grated manioc; fried manioc. Also known as *aipim*, *macaxeira*, *mandioca-doce*, *mandioca-mansa*. See also *bobó*.

mandioca(noun) manioc, cassava A tropical starchy tuberous root that is poisonous when raw.//Ex.: *Cooked, dried, then ground into flour or a coarse meal, it is one of the key ingredients, along with beans, fish, and rice, used in the cooking of Brazil.* [TCOB]//The hydrogen cyanide present in these tubers must be removed by extensive washing and cooking before they become edible. The juice extracted from the pulp is detoxified by boiling and becomes the basis for tucupi. Several kinds of flour originate from the dried pulp. A dish called *manicoba* is prepared with its leaves, called *maniva*. Also known as *mandioca-amarga*, *mandioca-brava*. See also *farinha de mandioca*, *manicoba*, *maniva*, *polvilho*, *puba*, *carimã*.

Figura 5

Modelo de verbete: 'mandioca'.

5) Análise

Conforme foi possível observar, a pesquisa terminográfica bilingue (assim como a monolíngue) nem sempre consegue alcançar a univocidade tão cara à Terminologia. A pluralidade de termos na língua de chegada para designar um mesmo conceito da língua de partida evidencia uma prática muito comum: na ausência de obras terminográficas bi- ou multilíngues da área em questão, o tradutor/redator, não raro, atua como terminólogo, uma vez que precisa cunhar um termo para resolver uma questão de tradução.

A pesquisa terminográfica bilingue muito se beneficia da análise de *corpora* comparáveis e paralelos para a extração de terminologia, assim como de equivalentes e definições na língua de chegada. No entanto, devemos considerar que os próprios tradutores e redatores desses textos também

sofrem com a escassez de obras terminográficas padronizadoras. Nos textos de culinária brasileira, observamos que tal lacuna leva esses profissionais à improvisação, disseminando a falta de padronização da terminologia específica da área. Ainda assim, as 'soluções' oferecidas vão além das disponibilizadas em obras de referência e podem contribuir para a compilação de um glossário especializado nesse domínio.

Com este trabalho, não temos a pretensão de eliminar a pluralidade de termos e equivalentes encontrados, mas sim mostrar para o consulente como utilizá-los para produzir textos coerentes e fluentes em língua inglesa.

6) Conclusões e encaminhamentos

Nesta pesquisa, procuramos mostrar os caminhos que estão sendo percorridos para a construção de um glossário de termos da culinária típica brasileira, utilizando, como metodologia, a Linguística de *Corpus*. Partimos da constatação de que, nessa área, não existem obras terminográficas no par de línguas português-inglês. A fim de preencher essa lacuna, utilizamos um *corpus* de estudo comparável em inglês e em português, construído para fins de levantamento de termos na LP (português) e de seus respectivos equivalentes, definições, colocações, fraseologias e outras informações na LC (inglês).

Tendo em vista que a obra em questão tem, como principal público-alvo, tradutores e redatores que visam transpor para a língua inglesa receitas, cardápios e outros textos que envolvam aspectos da culinária brasileira, concluímos que tal obra não pode se limitar a fornecer equivalentes em língua inglesa, mesmo porque nem sempre existe essa equivalência. Assim, as entradas que formarão o glossário contarão com definições, exemplos e outras informações, sempre tendo em vista o público-alvo.

A metodologia descrita tem-se mostrado eficaz para a construção do glossário de termos da culinária típica brasileira e acreditamos que possa ser aplicada, também, para a construção de glossários bilingües de outras áreas de especialidade.

Referências bibliográficas

- ALGRANTI, M. *Pequeno dicionário da gula*. Rio de Janeiro/São Paulo: Record, 2004.
- ALMEIDA, G. M. B.; CORREIA, M. Terminologia e corpus: relações, métodos e recursos. In: TAGNIN, S. E. O.; VALE, O. A. (Org.) *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008.
- AUBERT, F. H. *Introdução à metodologia da pesquisa bilingue*. São Paulo: Humanitas, FFLCH/USP, 1996.
- BACCIN, P. G. *Italianismos na terminologia gastronômica na cidade de São Paulo: um modelo de glossário*. 2003. 2v. Tese (Doutorado em Letras Modernas) — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2003.
- BARROS, L. A. *Curso básico de Terminologia*. São Paulo: Edusp, 2004.
- BERBER SARDINHA, A. P. *Linguística de corpus*. Barueri: Manole, 2004.
- BEVILACQUA, C. R. Unidades fraseológicas especializadas: novas perspectivas para sua identificação e tratamento. In: KRIEGER, M. da G.; MACIEL, A. M. B. *Temas de Terminologia*. Porto Alegre/São Paulo: Ed. Universidade/URGS/Humanitas/USP, 2001. p. 106-117.
- BOWKER, L.; PEARSON, J. *Working with specialized language: a practical guide to using corpora*. London/New York: Routledge, 2002.
- CAMARGO, D. C. de; VALIDÓRIO, V. C. Traduzindo elementos culturais em obras amadianas. In: X ENCONTRO NACIONAL DE TRADUTORES & IV ENCONTRO INTERNACIONAL DE TRADUTORES (ABRAPT), 2009, Ouro Preto. *Anais...*. Ouro Preto: UFOP, 2009. p. 1109-1122.
- CAPATTI, A.; MONTANARI, M. *La cucina italiana. Storia di una cultura*. Roma/Bari: Editori Laterza, 1999.
- CARLI, F.; KILOTZ, E. *Dicionário gastronômico*. São Paulo: Confronto, 2007.
- CORRÊA, R. H. M. A.; COSTA, E. G.; TAILLEFER, R. J. F.; ZORZO, V. F. Bahia com pimenta: um estudo comparado da tradução da culinária de D. Flor para o francês, o inglês e o espanhol. *Terra Roxa e Outras Terras* (3): 52-68, 2003. Disponível em: <http://www.uel.br/pos/letras/terraroxa/g_pdf/vol3/vol3_df.pdf>. Acesso em: 01. jul. de 2012.

- COSTA, A. T. P. *Brasil mostrando a sua cara: estratégias de tradução no material de divulgação cultural – um estudo baseado em corpus*. 262 f. Dissertação (Mestrado em Línguas Estrangeiras) — Universidade de Brasília, Brasília, 2006.
- DUBUC, R. *Manual prático de Terminologia*. Tradução Illeana Cabrera. Providência: Ril, 1999. Tradução de: Manuel pratique de terminologie.
- FERREIRA, A. B. de H. *Novo dicionário da língua portuguesa*. 2ª ed. Rio de Janeiro: Nova Fronteira, 1986.
- JAKOBSON, R. On Linguistic aspects of translation. 1959. In: VENTUTI, L. (Ed.) *The translation studies reader*. New York/London: Routledge, 2004.
- KLIE, V. *Glossário de gastronomia português-ínglês/ínglês-português*. Série Easy Way. São Paulo: Disal, 2006.
- KRIEGER, M. G.; M. J. B. FINATTO. *Introdução à Terminologia: teoria & prática*. São Paulo: Contexto, 2004.
- LÉVI-STRAUSS, C. The culinary triangle 1966. In: COUNIHAN, C. *Food and culture: a reader*. New York: Routledge, 2007, p. 28-35.
- MACIEL, A. M. B. Terminologia, linguagem de especialidade e dicionários. In: KRIEGER, M. G.; MACIEL, A. M. B. *Temas de Terminologia*. Porto Alegre/São Paulo: Ed. Universidade/UFRGS/Humanitas/USP, 2001, p. 39-46.
- MONTANARI, M. (Org.) *O mundo na cozinha. História, identidade, trocas*. São Paulo: Estação Liberdade/Senac, 2009.
- MORAES, H. F. R. de. *O advérbio em fraseologias das áreas da culinária e do direito contratual: um estudo baseado em corpus*. 448f. Tese (Doutorado em Letras Modernas) — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2010.
- OAKES, M. P. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press, 1998.
- ORENHA-OTTAIANO, A. Algumas contribuições advindas da compilação de corpora especializados via Web e WebBootCat para a tradução, terminologia e fraseologia. In: TAGNIN, S. E. O.; VALE, O. A. (Org.) *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008, p. 137-165.
- PEARSON, J. Como ter acesso a elementos definitórios nos textos especializados? Tradução: Carolina Huang e Sandra Dias Loguercio. In: *Cadernos de Tradução*, Porto Alegre, nº 17, outubro-dezembro, 2004, p. 51-66.
- PICHT, H. *Corpora como ponto de partida para a extração de dados terminológicos*. *Cadernos de Tradução*, Porto Alegre, nº 17, outubro-dezembro, 2004, p. 67-77.
- REBECHI, R. R. *A imagem do brasileiro no discurso do norte-americano em livros de culinária brasileira*. Dissertação (Mestrado em Letras Modernas — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2010a.
- REBECHI, R. R. The discourse of Americans in Brazilian cookbooks: a proposal of an analysis based on Corpus Linguistics. In: MOSKOVICH, I.; CRESPO, B.; LAREO, I.; LOJO, P. (Eds.). *LanguageWindowthroughCorpora / Visualización del lenguaje a través del corpus. II INTERNATIONAL CONFERENCE ON CORPUS LINGUISTICS (CLIC)*, 2010, A Coruña. *Proceedings...* A Coruña: Universidade da Coruña, 2010b, pp. 731-745.
- REBECHI, R. R. O que é culinária brasileira para o norte-americano? Um estudo baseado em Linguística de Corpus. In: PINTO, M. V.; SHEPHERD, T. M. G.; BERBER SARDINHA, T. (Org.). *Caminhos da Linguística de Corpus*. 1ed. Campinas: Mercado de Letras, 2012, v. 1, p. 375-404.
- ROCHA, M. Métodos estatísticos comuns em Linguística de Corpus: visão geral. In: GERBER, R. M.; VASILÉVSKI, V. (Org.) *Um percurso para pesquisas com base em corpus*. Florianópolis: Ed. da UFSC, 2007, p. 194-221.
- SCOTT, M. *Wordsmith Tools 6.0*. Oxford: Oxford University Press, 2012.
- TAGNIN, S. E. O.; TEIXEIRA, E. D. Linguística de Corpus e tradução técnica: relato da montagem de um corpus multivarietal de culinária. *TradTerm*, São Paulo, v. 10, 2004, p. 313-358.
- TAGNIN, S. E. O.; TEIXEIRA, E. D. *Vocabulário para culinária: inglês/português*. São Paulo: SBS, 2008.
- TAGNIN, S. E. O. *O jeito que a gente diz: combinações consagradas em inglês e português*. São Paulo: Disal, 2013.
- TEIXEIRA, E. D. *Receita qualquer um traduz. Será? – a culinária como área técnica de tradução*. Dissertação (Mestrado em Letras Modernas — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

TEIXEIRA, E. D. *A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual*. Tese (Doutorado em Letras Modernas) — Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2008.

ZILIO, L. Contraste alemão-português de fraseologias especializadas em textos de cardiologia. In: TAGNIN, S. E. O.; VALE, O. A. (Org.) *Avanços da Linguística de Corpus no Brasil*. São Paulo: Humanitas, 2008, p. 287-313.

Anexo

Livros que compõem o subcorpus em português

Título	Autoria	Ano	Sigla utilizada na pesquisa
Delícias da Cozinha Alagoana	Rocha et al.	1997	DCA
A Pátria nas Panelas	Cavalcanti	2007	PNP
Sabores da Cozinha Brasileira	Barbosa et al.	2004	SCB
Sabores e Saberes do Ceará	Rocha	2003	SSC
Revista Cláudia – Cozinha Brasileira	Sem data		SBC
Viagem Gastronômica Através do Brasil	Fernandes	2007	VGAB
Brasil: Gastronomia, Cultura e Turismo	Steinbruch	2010	BGCT
Culinária Brasileira, com Certeza	Campos	2012	CBCC

Livros que compõem o subcorpus em inglês

Título	Autor(es)	Ano de publicação	Sigla utilizada na pesquisa
A Taste of Brazil	Neelaman e Neelaman	2007	ATB
Brazilian Cooking	Leroux	1980	BC
Brazil: a Cook's Tour	Idone	1995	BCT
Brazil: a Culinary Journey	Hamilton	2005	BCJ
Cooking the Brazilian Way	Behnke e Duro	2004	CTBW
Delightful Brazilian Cooking	Ang	1993	DBC
Eat Smart in Brazil	Peterson e Peterson	1995	ESB
The Cooking of Brazil	Loerichio	2005	TCOB
The Brazilian Kitchen	Schwartz	2010	TBK
The Brazilian Table	Roberts e Roberts	2009	TBT
Brazilian Cookery	Andrade	2008	Braz.Cook.

CORPORA E VARIANTES CULTURAIS: UM ESTUDO DE CASO DA HOTELARIA

Sandra Navarro (USP)

ABSTRACT: This article describes how the use of corpora in terminological research can shed light on cultural aspects of a specialized field. To that aim, we rely on data collected for our masters' research, which investigated the phraseological units typical of the hotel industry so as to create a bilingual glossary (English-Portuguese). The study was based on a bilingual corpus containing about 1 million words and 1,000 texts taken from websites of hotels in Brazil and the United States. Each step of the work with this corpus revealed aspects typical of the cultures represented therein. While planning the structure, compiling the texts, balancing the corpus and finally investigating the language, it was possible to retrieve information regarding, for instance, the categorization of the hotel field, the text typology of hotel websites and expectations of their readers as well as the existence of different practices and services in Brazil and the US. We believe that the main advantage of expanding the scope of linguistic research into the cultural dimension is the possibility of developing more comprehensive and detailed terminological materials.

KEYWORDS: Corpus Linguistics; Terminology; Hotel Industry; Culture; Phraseology

1) Introdução

Duas línguas não são nunca suficientemente semelhantes para serem consideradas como representantes da mesma realidade social. Os mundos nos quais as diferentes sociedades vivem são mundos distintos, e não simplesmente o mesmo mundo com rótulos diferentes (SAPIR, 1956, p. 69).

O uso de *corpora* no trabalho terminológico é prática já bastante difundida e imprescindível nos dias atuais. Suas vantagens também são amplamente conhecidas, como, por exemplo, maior facilidade, objetividade e confiabilidade na identificação de padrões lexicais. No entanto, em nossa pesquisa constatamos que o trabalho com *corpus* pode fornecer resultados que